

Universidade de Lisboa
Faculdade de Ciências
Departamento de Estatística e Investigação Operacional



**Modelos de sobrevivência para estudo do tempo até à
ocorrência de excesso de peso em indivíduos adultos
submetidos a Transplante Alogénico de Células Progenitoras
Hematopoiéticas**

Sónia Denise Ferreira Velho

Dissertação

Mestrado em Bioestatística

2015

Universidade de Lisboa
Faculdade de Ciências
Departamento de Estatística e Investigação Operacional



**Modelos de sobrevivência para estudo do tempo até à
ocorrência de excesso de peso em indivíduos adultos
submetidos a Transplante Alogénico de Células Progenitoras
Hematopoiéticas**

Sónia Denise Ferreira Velho

Dissertação orientada pela Prof^a Doutora Cristina Simões Rocha

Mestrado em Bioestatística

2015

Resumo

A obesidade e o excesso de peso é um dos principais problemas de saúde pública em Portugal, dada a sua crescente prevalência e comorbilidades. Existe evidência científica que mostra que fatores associados às doenças hematológicas malignas, podem predispor para o aumento do peso. Contudo, há poucos estudos que tenham investigado esta temática, e a maioria incidiu sobre doentes pediátricos. Neste estudo foram incluídos doentes adultos submetidos a Transplante de Células Progenitoras Hematopoiéticas Alogénico (TCPH-alo), dada a importância da obtenção de mais informação. Esta tese teve como objetivos efetuar a análise de sobrevivência sendo o acontecimento de interesse o aumento no Índice de Massa Corporal (IMC) para uma categoria superior de excesso de peso/obesidade após TCPH-alo; e identificar fatores que influenciem significativamente o tempo até à ocorrência do acontecimento de interesse. Os dados foram obtidos a partir do processo clínico de doentes transplantados no Instituto Português de Oncologia de Lisboa-Francisco Gentil, entre Maio de 1987 e Dezembro de 2007. A análise estatística foi feita na quase totalidade com o *software* R versão 3.0.2. Na análise preliminar foi utilizado o estimador de Kaplan-Meier e os testes log-rank, Gehan e Tarone-Ware. No estudo da influência das covariáveis no tempo até o acontecimento de interesse procedeu-se à comparação do modelo de Cox com modelos paramétricos.

Nas análises com os testes não paramétricos não se verificaram diferenças estatisticamente significativas, com a exceção da covariável IMC no TCPH-alo. Tanto no modelo semi-paramétrico de Cox como nos modelos paramétricos não foram encontradas covariáveis com influência no tempo até ao acontecimento de interesse.

Com base nos dados obtidos observou-se que uma elevada percentagem de doentes desenvolvem excesso de peso/obesidade após TCPH-alo. No entanto, o TCPH-alo não parece estar associado de forma significativa à etiologia do excesso de peso/obesidade, pelo que deverá ser dada maior atenção aos fatores inerentes ao padrão alimentar e estilo de vida.

Palavras-chave: Análise de sobrevivência, modelo de Cox, modelos paramétricos, obesidade, excesso de peso, doenças hematológicas malignas, TCPH-alo.

Abstract

Obesity and overweight are major public health concerns in Portugal, given its increasing prevalence and comorbidities. There is some scientific evidence that shows that malignant hematological diseases are associated to several factors that may predispose to weight gain. However, few studies have investigated this issue, and most of these studies have focused mainly on pediatric patients. This study included adult patients that were treated with Allogeneic Hematopoietic Stem Cell Transplantation (HSCT-alo), given the importance of obtaining information for this type of patients.

This study aimed at using survival analysis for the analysis of the increase in Body Mass Index (BMI) that results in a higher weight classification, such as obesity or overweight after HSCT-alo; and to identify factors that may influence time until the occurrence of the event of interest. Data was collected retrospectively from medical files of transplanted patients at the Portuguese Institute of Oncology of Lisbon-Francisco Gentil. Statistical analysis was performed using mainly the statistical software R version 3.0.2. Preliminary analysis was conducted using Kaplan-Meier estimator and log-rank, Gehan and Tarone-Ware tests.

The influence of covariates on the time until occurrence of the event of interest was analysed with both Cox model and parametric models. Non-parametric tests resulted in no statistically significant covariates, with the exception of BMI during HSCT-alo. No covariates were found to influence significantly the time until the event of interest in both semiparametric model and parametric models.

Based on the data from this study, a high percentage of patients develop overweight or obesity after HSCT-alo. However, HSCT-alo does not appear to be significantly associated with the etiology of overweight/obesity and greater attention should be given to other factors such as dietary pattern and lifestyle.

Key-words: Survival analysis, Cox model, parametric models, obesity, overweight, haematological malignancies, HSCT-alo.

Índice	
Resumo.....	i
Abstract	ii
Índice de tabelas.....	vi
Índice de figuras.....	vii
Agradecimentos.....	ix
Lista de abreviaturas.....	x
Glossário	xi
Capítulo1- O Transplante de Células Progenitoras Hematopoiéticas e o Excesso de Peso e Obesidade	1
1.1 Introdução	1
1.2 Objetivos	2
1.3 Enquadramento histórico do Transplante de Células Progenitoras Hematopoiéticas	2
1.4 O que é a medula óssea?	3
1.5 Tipos de Transplante de Células Progenitoras Hematopoiéticas.....	4
1.5.1 Transplante de Células Progenitoras Hematopoiéticas Alogénico	4
1.5.2 Transplante de Células Progenitoras Hematopoiéticas Autólogo	5
1.6 Complicações do Transplante de Células Progenitoras Hematopoiéticas com impacto na alimentação e no estado nutricional	6
1.7 A obesidade e o risco de doenças hematológicas malignas	7
1.8 Tratamento de doenças hematológicas malignas e o risco de excesso de peso/obesidade.....	8
Capítulo 2-Análise de Sobrevivência.....	11
2.3 Inferência estatística não paramétrica	14
2.3.1 Estimador da função de sobrevivência.....	14
2.3.2 Estimador da função de risco cumulativa.....	16
2.3.3 Testes para comparação de curvas de sobrevivência	17
2.4 Modelo de Cox.....	23
2.4.1 Função de verosimilhança parcial	24
2.4.2 Existência de observações empatadas	25
2.4.3 Intervalos de confiança e testes de hipóteses para β	25
2.4.4 Estimação da função de sobrevivência.....	26

2.4.5	Comparação de distribuições do tempo de vida.....	27
2.4.6	Método para a seleção de covariáveis	28
2.4.7	Comparação de modelos alternativos.....	29
2.4.8	Métodos de diagnóstico para o modelo de Cox	31
2.5	Modelos de sobrevivência paramétricos.....	33
2.5.1	Distribuição exponencial.....	33
2.5.2	Distribuição de Weibull.....	34
2.5.3	Distribuição log-logística	35
2.5.4	Distribuição log-normal	36
2.5.5	Distribuição Gama	36
2.5.6	Distribuição de Gompertz	37
2.5.7	Função de verossimilhança.....	38
2.6	Modelos de regressão paramétricos.....	38
2.6.1	Modelos de tempo de vida acelerado.....	39
2.6.2	Modelo de riscos proporcionais.....	40
2.6.3	Modelos de possibilidades proporcionais	41
2.6.4	Modelo de regressão de Weibull	41
2.6.5	Modelo de regressão log-logístico	42
2.6.6	Critério de Informação de Akaike.....	43
2.6.7	Análise de Resíduos	44
Capítulo 3-Análise do tempo até à ocorrência de excesso de peso/obesidade		47
3.1	Descrição do estudo	47
3.2	Definição das variáveis.....	48
3.3	Análise preliminar.....	50
3.4	Estimação não paramétrica.....	54
3.5	Comparação entre curvas de sobrevivência	58
3.6	Modelo de Cox.....	60
3.7	Análise de resíduos para o modelo de Cox.....	63
3.8	Modelos de regressão paramétricos.....	67
3.9	Análise de resíduos para os modelos paramétricos.....	70
3.10	Comparação gráfica de modelos	75
Capítulo 4-Resultados		77

4.1	Modelo de Cox.....	77
4.2	Modelos de regressão paramétricos	78
4.3	Discussão e Conclusão	80
	Bibliografia.....	83
	Anexo I.....	85
	Anexo II.....	87

Índice de tabelas

Tabela 1: Comparação de modelos paramétricos.....	44
Tabela 2: Definição das variáveis.....	48
Tabela 3: Definição das variáveis (continuação).	49
Tabela 4: Caracterização da amostra estudada.	50
Tabela 5: Caracterização da amostra estudada (continuação).....	51
Tabela 6: Caracterização da evolução ponderal.....	54
Tabela 7: Testes não paramétricos log-rank, Gehan e Tarone-Ware.....	59
Tabela 8: Estatística $-2\log\hat{L}$ e respetivo valor-p para cada modelo ajustado.	62
Tabela 9: Resultados do teste de correlação entre os resíduos de Schoenfeld	64
Tabela 10: Estimativas dos parâmetros e os valores de AIC para os modelos de regressão exponencial, Weibull, log-logístico e log-normal	68
Tabela 11: Estimativas dos parâmetros e intervalos de 95% de confiança obtidos a partir do modelo de Cox.	77
Tabela 12: Estimativas dos parâmetros obtidos a partir do modelo de Cox estratificado para o IMC no transplante e com inclusão da covariável diagnóstico.	78
Tabela 13: Estimativas dos parâmetros obtidos a partir do modelo Weibull na forma de modelo de tempo de vida acelerado.	78
Tabela 14: Estimativas dos parâmetros obtidas a partir do modelo Weibull na forma de modelo de riscos proporcionais e para o modelo de Cox.	79

Índice de figuras

Figura 1: Medula óssea	3
Figura 2: Transplante de Células Progenitoras Hematopoiéticas alogénico	5
Figura 3: Transplante de Células Progenitoras Hematopoiéticas autólogo	6
Figura 4: Distribuição do peso por género em quatro momentos de avaliação	52
Figura 5: Histograma da variável peso nos diferentes momentos de avaliação.	53
Figura 6: Estimativa de Kaplan-Meier da função de sobrevivência e intervalos de confiança com transformação Log-Log.....	55
Figura 7: Estimativa de Kaplan-Meier da função de sobrevivência obtida a partir do estimador de para as variáveis: Diagnóstico, Género, Idade no transplante, IMC no transplante, TBI, Ciclosporina, Mofetil, Tacrolimus, Corticoide, GVHD, Hipertensão, Dislipidemia e Diabetes.....	58
Figura 8: Comparação de estimativas da função de risco.	58
Figura 9: Gráficos $\log(-\log\hat{S}(t))$ em função do tempo para as variáveis Idade no transplante categorizada.	60
Figura 10: Gráficos $\log(-\log\hat{S}(t))$ em função do tempo para as variáveis Diagnóstico, IMC no transplante categorizada e TBI.....	61
Figura 11: Estimativa da função de risco cumulativa para os resíduos <i>versus</i> os resíduos de Cox-Snell.	63
Figura 12: Análise da premissa de riscos proporcionais para a covariável diagnóstico e IMC categorizada com os resíduos de Schoenfeld padronizados <i>versus</i> tempo transformado.....	64
Figura 13: Resíduos Martingala do modelo ajustado <i>versus</i> índice do indivíduo.	65
Figura 14: Desvios residuais <i>versus</i> índice.	66
Figura 15: Desvios residuais <i>versus</i> o tempo.	66
Figura 16: Resíduos score para as covariáveis IMC no transplante e Diagnóstico.	67
Figura 17: Estimativa da função de risco cumulativa para os resíduos <i>versus</i> os resíduos de Cox-Snell.	71
Figura 18: Resíduos Martingala do modelo ajustado <i>versus</i> índice do indivíduo.....	71
Figura 19: Desvios residuais <i>versus</i> índice.	72
Figura 20: Desvios residuais <i>versus</i> tempo.	72
Figura 21: Estimativa da função de risco cumulativa <i>versus</i> os resíduos de Cox-Snell. .	73

Figura 22: Resíduos Martingala do modelo ajustado <i>versus</i> índice do indivíduo.....	73
Figura 23: Desvios residuais <i>versus</i> índice.	74
Figura 24: Desvios residuais <i>versus</i> índice.	74
Figura 25: Estimativas da função de sobrevivência obtidas com modelos paramétricos (exponencial e Weibull) e com o estimador de Kaplan-Meier.....	75

Agradecimentos

À Professora Doutora Cristina Simões Rocha orientadora desta tese de Mestrado, por toda a sua disponibilidade e capacidade de ensino de conceitos tão distantes da minha formação de base. A concretização desta tese foi um verdadeiro desafio, tendo permitido um desenvolvimento pessoal importante que julgo que não teria sido possível sem a orientação e o apoio experienciado.

Ao Professor Doutor Manuel Abecassis, Diretor da Unidade de Transplante de Medula do Instituto Português de Oncologia de Lisboa-Francisco Gentil, por toda a atenção e facilidades concedidas na obtenção dos dados.

Aos Profissionais de Saúde da Unidade de Transplante de Medula do Instituto Português de Oncologia de Lisboa-Francisco Gentil, e em particular à Dr.^a Isabelina Ferreira pelos preciosos comentários e ideias que geraram a necessidade deste estudo; à Enfermeira Chefe Elsa Oliveira pela boa disposição e interesse e à Enfermeira Rosália Pires por todo o apoio e compreensão.

Aos colegas da Unidade Autónoma de Nutrição e Dietética do Instituto Português de Oncologia de Lisboa-Francisco Gentil, em especial à Dr.^aEugénia Santos Silva pela amizade e apoio.

Aos Doentes com doença hematológica maligna pela simpatia, força e determinação que sempre me transmitiram.

À minha mãe pelo apoio incondicional e por todas as palavras sábias no momento certo, que me fizeram acreditar que tudo na vida é possível. A concretização desta tese é mais uma prova disso.

À minha irmã e ao meu pai por toda a força, persistência e capacidade de trabalho que me inculcaram.

Aos meus amigos que me ajudaram a descontrair em momentos críticos da realização desta tese de Mestrado.

Por fim, um agradecimento muito especial ao Carlos Vilar pelo apoio e atitude positiva que me ajudaram a dissolver todos os momentos difíceis, e a manter-me concentrada nos meus objetivos pessoais. Sei que sou melhor ao teu lado.

Lista de abreviaturas

CPH- Células Progenitoras Hematopoieticas

GVHD- *Graft vs. Host Disease*

HC- Hormona de Crescimento

HTA- Hipertensão Arterial

IC- Intervalo de Confiança

IMC- Índice de Massa Corporal

IMCT- Índice de Massa Corporal no Transplante

K-M- Estimador de Kaplan-Meier

LBDGC- Linfoma B de Grandes Células

LLA- Leucemia Linfóide Aguda

LLC- Leucemia Linfóide Crónica

LMA- Leucemia Mielóide Aguda

LMC- Leucemia Mielóide Crónica

LNH- Linfoma Não Hodgkin

N-A- Estimador de Nelson-Aalen

SM- Síndrome Metabólico

TBI- *Total Body Irradiation*

TCPH-a- Transplante de Células Progenitoras Hemtopoiéticas Autólogo

TCPH-alo- Transplante de Células Progenitoras Hematopoiéticas Alogénico

Glossário

Anorexia-falta de apetite.

Disgeusia- alteração do paladar.

Leucemia refratária- leucemia que não entra em remissão após o tratamento

Mielosupressão- supressão da produção de células pela medula

Mucosite- inflamação da mucosa que pode atingir a totalidade do tubo digestivo.

Neutropénia- redução da contagem dos linfócitos que pode ser causada pela quimio e radioterapia, e que provoca uma maior suscetibilidade para infeções.

Nutrição parentérica-fornecimento de nutrientes na veia periférica ou central através de um cateter venoso central.

Odinofagia-dor na mastigação e deglutição.

Pancitopénia- redução de todas as séries celulares do sangue.

Remissão- período de ausência de doença

Xerostomia- boca seca.

Capítulo1- O Transplante de Células Progenitoras Hematopoiéticas e o Excesso de Peso e Obesidade

1.1 Introdução

A bioestatística é a aplicação da estatística ao campo biológico e médico. Esta ciência é reconhecida como uma ferramenta fundamental para a análise e interpretação de dados, com vista à obtenção de conclusões fundamentadas. Um adequado domínio da bioestatística permite uma estruturação rigorosa da metodologia de investigação, que é crucial para garantir a qualidade dos resultados obtidos. A bioestatística aplicada à Nutrição é atualmente uma necessidade, uma vez que se trata de uma área da saúde em plena expansão, cujos progressos científicos poderão ter repercussões significativas na saúde e qualidade de vida da população. Efetivamente, a obesidade é atualmente um dos principais problemas de saúde pública em Portugal, dada a sua crescente prevalência e comorbilidades associadas.

Existe uma escassez de estudos sobre a obesidade e o excesso de peso em doentes oncológicos, sendo apenas um pequeno número destes estudos dedicado às patologias hematológicas malignas. Para além disto, é importante salientar que a grande maioria dos estudos que investigaram a obesidade e o excesso de peso em sobreviventes de doenças hematológicas malignas incidiram sobre indivíduos que manifestaram a doença em idade pediátrica. Assim sendo, o excesso de peso e obesidade em sobreviventes de doenças hematológicas malignas que foram submetidos a transplante de células progenitoras hematopoiéticas em idade adulta foi a temática escolhida para o presente estudo, dado o grande interesse da obtenção de resultados para este tipo de doentes.

1.2 Objetivos

Objetivos gerais

- Efetuar a análise de sobrevivência utilizando como o acontecimento de interesse a alteração do Índice de Massa Corporal para uma categoria superior correspondente a excesso de peso ou obesidade após Transplante de Células Progenitoras Hematopoiéticas Alogénico (TCPH-alo).
- Identificar fatores que influenciem significativamente o tempo até à ocorrência do acontecimento de interesse.

Para atingir estes objetivos procedeu-se a uma:

- Análise preliminar dos dados com a utilização de métodos não paramétricos (estimador de Kaplan - Meier, testes log-rank, Gehan e de Tarone-Ware).
- Comparação do modelo de regressão de Cox com modelos de regressão paramétricos (modelo Weibull e exponencial) no estudo da influência das covariáveis no tempo até ao aumento no IMC com passagem para uma categoria de excesso de peso ou obesidade.

A análise estatística foi feita na quase totalidade com o *software* estatístico R versão 3.0.2.

1.3 Enquadramento histórico do Transplante de Células Progenitoras Hematopoiéticas

A primeira transplantação de medula óssea foi efetuada no final da década de 60 do século XX em 3 doentes com imunodeficiência congénita. No início dos anos 70, o Dr. Edward Donnall Thomas e colegas demonstraram que o transplante de medula levava a uma maior sobrevivência em doentes com diagnóstico de leucemia refratária. Posteriormente foi demonstrado um melhor resultado clínico associado ao transplante na fase inicial da doença. Na década de 80, o transplante de medula alogénico passou a ser amplamente aceite para o tratamento da imunodeficiência congénita, anemia aplástica, leucemia aguda e crónica. Nesta década foi também dado início ao transplante de medula alogénico a partir de dador não relacionado. No final desta década realiza-se o primeiro transplante de células progenitoras hematopoiéticas obtidas a partir do cordão umbilical. Em 1990, o Dr. Edward Donnall Thomas é reconhecido com o Prémio Nobel da Medicina (Tan, 2004).

Em Portugal, a introdução dos transplantes de medula óssea foi feita em 1987, pelo Prof. Manuel Abecassis no Instituto Português de Oncologia de Lisboa - Francisco Gentil (IPOL-FG). Em Maio de 1989, o Hospital de Santa Maria passou também a realizar transplantes de medula óssea e, no mesmo ano, realiza-se o primeiro transplante de medula no Instituto Português de Oncologia do Porto.

É de destacar que na Unidade de Transplante de Medula Óssea do IPOL-FG ao longo dos anos foram implementadas metodologias de transplante pioneiras, nomeadamente a primeira transplantação de células do cordão umbilical (1994), os primeiros transplantes de medula alogénicos com células do sangue periférico, a primeira transplantação com dador não relacionado (1995) e a primeira transplantação em ambatório (1999).

Por último, é ainda de referir que devido ao empenho do CEDACE (Centro Nacional de Dadores de Medula Óssea Estaminais ou de Sangue do Cordão) e dos Centros de Histocompatibilidade do norte, do centro e do sul, o registo português é atualmente o 2º maior registo da Europa e 3º do mundo. Este trabalho é fundamental dado que Portugal regista anualmente 1000 novos casos de leucemia, e uma elevada percentagem destes doentes irá necessitar de realizar um transplante de medula óssea como recurso terapêutico final para assegurar a sobrevivência.

1.4 O que é a medula óssea?

A medula óssea é um tecido esponjoso que preenche o interior de vários ossos, como por exemplo os da bacia. É neste tecido que existem as células progenitoras hematopoiéticas (CPH), que são células imaturas com capacidade para se diferenciarem em qualquer célula do sangue periférico, nomeadamente as hemácias, plaquetas e leucócitos (figura 1). O processo de formação das células sanguíneas é chamado hematopoiese.

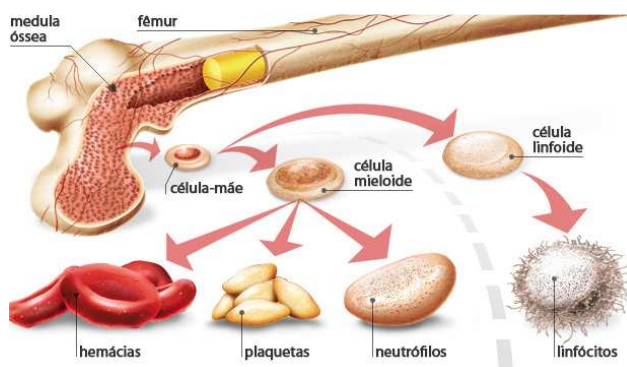


Figura 1: Medula óssea (www.apcl.pt).

1.5 Tipos de Transplante de Células Progenitoras Hematopoiéticas

Nos últimos anos, o termo “transplante de medula óssea” foi substituído por “transplante de células progenitoras hematopoiéticas”(TCPH), visto ser um termo mais preciso, por atribuir maior ênfase às células progenitoras hematopoiéticas (CPH). O TCPH está indicado no tratamento de doenças hematológicas malignas (leucemia e linfoma), tumores sólidos e distúrbios autoimunes.

As CPH são células imaturas, com elevado potencial de renovação e capacidade de povoar a medula óssea. Estas células podem ser extraídas da medula óssea, sangue periférico ou do cordão umbilical. Nas últimas duas décadas, as células progenitoras hematopoiéticas (CPH) têm sido principalmente colhidas no sangue periférico, após a estimulação da medula óssea com fatores de crescimento. Contudo, a medula óssea e as células do cordão umbilical de um dador relacionado ou não relacionado também têm sido usadas no tratamento de doenças hematológicas.

O TCPH designa-se como autólogo, alogénico ou singénico consoante as células progenitoras hematopoiéticas sejam oriundas do próprio doente, de um dador (familiar ou não familiar histocompatível) ou de um gémeo idêntico, respetivamente.

1.5.1 Transplante de Células Progenitoras Hematopoiéticas Alogénico

No caso do Transplante de Células Progenitoras Hematopoiéticas Alogénico (TCPH-alo), o regime de condicionamento é por norma mais agressivo, visto ser administrado concomitantemente quimioterapia em elevada dose e radioterapia corporal total (*Total Body Irradiation*-TBI). O regime de condicionamento tem por finalidade promover a imunossupressão para evitar a destruição do enxerto por células imunologicamente ativas residuais do hospedeiro, eliminar eventuais células cancerígenas e criar espaço para o desenvolvimento do novo sistema imunitário. Alguns dos citotóxicos mais utilizados no regime de condicionamento incluem agentes alquilantes (ciclofosfamida e bussulfano) e antimetabólicos (metotrexato e fludrabin). Esta medicação pode provocar sintomas com impacto nutricional, nomeadamente a mielossupressão, anorexia, náuseas, vômitos e fadiga. Em particular, os agentes alquilantes podem igualmente provocar toxicidade renal e os agentes antimetabólicos, diarreia e mucosite (Mahan e Escott-Stump, 2010).

Uma das principais complicações do TCPH-alo é a Doença do Enxerto Contra o Hospedeiro (*Graft vs. Host Disease*-GVHD), que resulta da ação das células imunocompetentes do enxerto contra o hospedeiro, podendo ter envolvimento cutâneo, hepático ou gastrointestinal. Esta complicação manifesta-se normalmente

nos primeiros 100 dias após o transplante, mas também pode ser observada após 7-10 dias. A GVHD pode ser aguda ou crônica, e neste caso desenvolver-se até 3 meses depois do transplante, sendo mais frequentemente observado quando os doadores são não idênticos e não familiares.

Assim, após o TCPH é utilizada medicação imunossupressora como profilaxia da GVHD, que inclui ciclosporina, tacrolimus, corticóide e micofenolato de mofetil. Os corticóides apresentam como principais efeitos secundários com influência no estado nutricional a retenção de líquidos e de sódio, intolerância à glucose, espoliação de potássio e aumento do apetite.

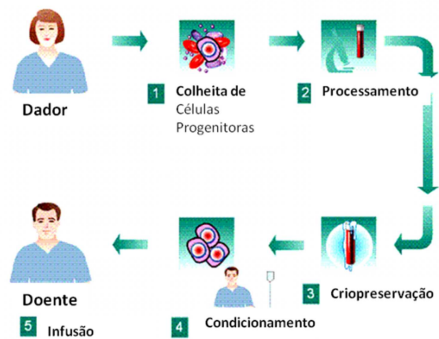


Figura 2:Transplante de Células Progenitoras Hematopoiéticas alogénico (TCH-a)
(www.lymphomation.org).

1.5.2 Transplante de Células Progenitoras Hematopoiéticas Autólogo

O TCH-a consiste na administração de CPH do próprio doente, após condicionamento com quimioterapia de elevada dose. Neste caso, não existe o risco de GVHD, o que implica à partida menor morbilidade e mortalidade e custos inferiores associados ao tratamento destes doentes.

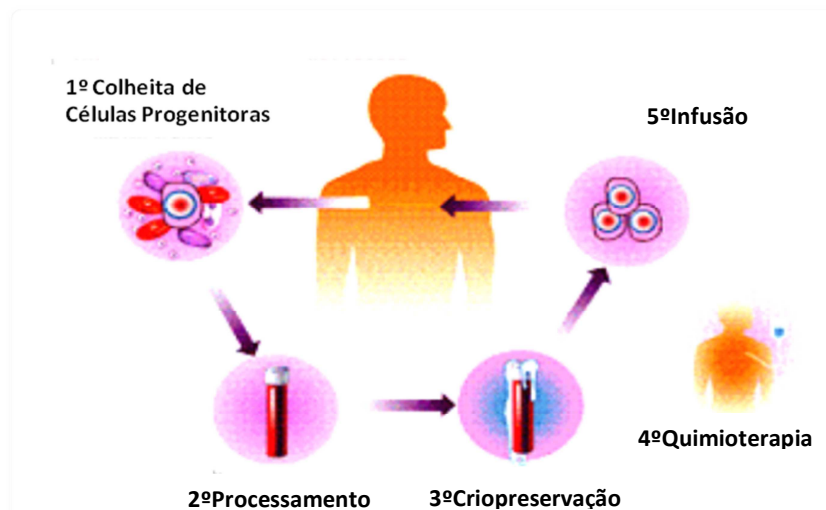


Figura 3: Transplante de Células Progenitoras Hematopoiéticas autólogo (TPCH-a) (www.lymphomation.org).

1.6 Complicações do Transplante de Células Progenitoras Hematopoiéticas com impacto na alimentação e no estado nutricional

Em ambos os tipos de transplante, o regime de condicionamento vai induzir a mielossupressão e provocar pancitopenia. Durante este período, o doente não apresenta um sistema imunitário competente, e como tal está mais suscetível a infeções. Por este motivo, são implementadas restrições alimentares e redobrados os cuidados de higiene na preparação e manipulação dos géneros alimentícios, de modo a fornecer uma alimentação de baixa carga bacteriana, com a finalidade de evitar infeções veiculadas pelos alimentos.

O TCPH tem consequências nutricionais graves, tais como graus variáveis de mucosite, xerostomia e disgeusia. A mucosite é em geral dolorosa e desenvolve-se em cerca de 75% dos casos. Durante o TCPH, a consistência da dieta fornecida é ajustada à capacidade de mastigação e deglutição do doente, verificando-se que de um modo geral os alimentos líquidos e macios são melhor tolerados. Nalguns casos de mucosite grave a alimentação por via oral pode estar impossibilitada, sendo necessário o recurso à nutrição parentérica.

No TCPH-a, o período de pancitopénia tem vindo a diminuir devido ao recurso a fatores de crescimento e assim também se verificou uma redução no número de dias de mucosite. Em contrapartida, o regime de condicionamento administrado no TCPH-alo está associado a uma maior toxicidade gastrointestinal, caracterizada por mucosite, odinofagia e/ou diarreia, náuseas e vómitos. O GVHD é uma das principais complicações do TCPH-alo, que pode complicar ainda mais o estado nutricional do

doente. O GVHD agudo com envolvimento gastrointestinal é uma situação grave, na qual o doente pode apresentar um quadro de gastroenterite, que se caracteriza por dor abdominal, náuseas e vômitos. O GVHD crónico pode afetar a pele, a mucosa oral (estomatites, ulcerações, xerostomia) e o trato gastrointestinal (anorexia, sintomas de refluxo e diarreia). A doença veno-oclusiva é outra complicação grave e por vezes fatal, que pode ocorrer tanto no TCPH-a como no TCPH-alo, como consequência da toxicidade da quimioterapia.

De um modo global, as complicações do TCPH predisõem para a desnutrição, sendo por isso necessária a intervenção nutricional nestes doentes. Existe alguma evidência científica que mostra que o baixo peso durante o TCPH é um indicador de mau prognóstico (Muscaritoli, et al, 2002). Em contrapartida, alguns estudos efetuados em sobreviventes de TCPH de idade pediátrica, têm mostrado que após o TCPH ocorre uma diminuição da massa magra e um aumento da percentagem de massa gorda, sendo por isso provável um maior risco de comorbilidades associadas ao excesso de peso/obesidade (Kyle, et al., 2005).

1.7 A obesidade e o risco de doenças hematológicas malignas

A obesidade caracteriza-se pelo excesso de tecido adiposo, e é atualmente aceite a definição com base no cálculo do Índice de Massa Corporal (IMC), que é dado por: $\text{Peso (kg)} / \text{Altura (m)}^2$. De acordo com a Organização Mundial de Saúde indivíduos com IMC entre 25 e 29.9 apresentam excesso de peso, sendo um IMC superior a 30 correspondente a obesidade. A obesidade pode ainda ser classificada em tipo I (IMC:30-34.9), tipo II (IMC:35-39.9) ou tipo III (IMC \geq 40) (WHO).

Um estudo abrangendo o período entre 2003-2005 mostrou que aproximadamente 39.4% da população portuguesa apresentava excesso de peso e 14.2% tinha obesidade (Carmo, et al., 2008). O excesso de peso/obesidade estão associados a diversas comorbilidades tais como: diabetes, hipertensão, dislipidemia, doença cardiovascular e certos tipos de cancro, nomeadamente cancro da mama, do endométrio, dos ovários, do cólon e da próstata. O mecanismo fisiológico subjacente à associação entre o excesso de peso/obesidade e cancro assenta na interação entre vias metabólicas e hormonais, na qual a resistência à insulina assume um papel central (Ceschi, et al., 2007). No entanto, poucos estudos têm sido efetuados que permitam perceber se existe também uma associação entre o excesso de peso/obesidade e as doenças hematológicas malignas. Em particular, alguns estudos têm evidenciado uma possível relação entre a obesidade/excesso de peso e o risco de certos tipos de leucemia e linfoma.

1.8 Tratamento de doenças hematológicas malignas e o risco de excesso de peso/obesidade

Durante o tratamento de doenças hematológicas malignas, diversos fatores parecem predispor para o desenvolvimento de excesso de peso/obesidade. Especificamente no tratamento da Leucemia Linfocítica Aguda (LLA), o aumento de peso e a alteração da composição corporal resulta da administração de glucocorticóides (prednisolona ou dexametasona). No entanto, o aumento de peso acelerado também tem sido documentado nos primeiros 2 anos após a suspensão do tratamento. Para além dos sobreviventes de LLA, a alteração na composição corporal, caracterizada pelo aumento de massa gorda foi observada em sobreviventes de outros tipos de Leucemia e Linfoma em idade pediátrica. Na população pediátrica é ainda comum o atraso no crescimento, como consequência de uma insuficiência parcial e transitória ou diminuição da sensibilidade à Hormona de Crescimento (HC). Esta disfunção endócrina é frequente após radioterapia craniana. Assim sendo, é esperado que um número considerável de sobreviventes de LLA poderá vir a desenvolver excesso de peso/obesidade (Siviero-Miachon, et al., 2009).

É de salientar que a HC, para além de estar implicada no crescimento, pode igualmente afetar o metabolismo. A deficiência em HC tem sido associada às comorbilidades implicadas na Síndrome Metabólica (SM) (Haas, et al., 2010). A SM é definida por um conjunto de fatores de risco cardiovasculares que incluem a obesidade abdominal, dislipidemia, hipertensão, resistência à insulina e um perfil pro-inflamatório e pro-trombótico. Alguns estudos têm mostrado que adultos e crianças sobreviventes de cancro apresentam um risco acrescido de virem a desenvolver resistência à insulina, SM e mortalidade associada a doença cardiovascular. Alguns estudos efetuados em adultos sujeitos a TCPH-allo mostraram que estes doentes apresentam um maior risco de diabetes, hipertensão e doença arterial vascular prematura. Pensa-se que os doentes sujeitos a TCPH-allo têm maior predisposição para o desenvolvimento de SM, devido a diversos mecanismos, tais como a lesão do sistema neurohormonal e do endotélio vascular, os efeitos imunitários e inflamatórios decorrentes do enxerto alogénico, a GVHD e a terapia associada. Contudo, os fatores de risco cardiovasculares em doentes que foram submetidos a TCPH-allo ainda não foram devidamente estudados. Num estudo transversal, caso controlo realizado em 86 adultos sujeitos a TCPH e 258 controlos emparelhados para a idade e género, constatou-se que a síndrome metabólica (SM) atingia 49% dos doentes transplantados e observou-se nestes doentes um risco de SM 2.2 vezes superior, relativamente ao grupo de controlo (Majhail, et al., 2009).

As causas para o potencial aumento na prevalência de obesidade/excesso após TCPH permanecem pouco claras. O aumento de peso excessivo durante o tratamento

tem sido maioritariamente associado aos efeitos secundários dos corticóides e à desregulação do apetite, assim como a um dispêndio energético mais baixo atribuível à inatividade física. No entanto, estes fatores não explicam por completo o excesso de peso neste grupo de doentes e a influência dos tratamentos com glucocorticoides e citoestáticos na composição corporal permanece incerta. Outros fatores que podem explicar o aumento de peso em sobreviventes de LLA têm a ver com o contexto familiar, a inadequada seleção de alimentos que integram o padrão alimentar, o sedentarismo e a reduzida atividade física.

A prevalência e a etiologia do excesso de peso/obesidade em sobreviventes de cancro, a sua relevância e a relação com o desenvolvimento de doença cardiovascular são áreas de investigação ainda pouco exploradas. Esta problemática necessita de ser esclarecida, visto existir evidência científica de que a obesidade está associada a uma maior mortalidade (por recidiva ou sem associação com o diagnóstico de base) e a uma maior taxa de complicações infecciosas no período pós-transplante (Fuji, et al., 2009).

Capítulo 2-Análise de Sobrevivência

Neste capítulo serão apresentados os conceitos básicos, métodos e modelos de análise de sobrevivência utilizados neste trabalho, seguindo principalmente (Collett, 2003) e (Rocha e Papoila, 2009).

Seja T uma variável aleatória, não negativa, absolutamente contínua que representa o tempo de vida de um indivíduo pertencente a uma população homogênea, isto é, um indivíduo que não difere dos restantes relativamente a fatores com potencial para influenciar a sua sobrevivência. A função densidade de probabilidade é dada por:

$$f(t) = \lim_{dt \rightarrow 0^+} \frac{P(t \leq T \leq t + dt)}{dt}$$

A função de sobrevivência no instante t , define-se como a probabilidade de um indivíduo sobreviver para além do instante t e é dada por:

$$S(t) = P(T > t), t \geq 0$$

Trata-se de uma função monótona decrescente e contínua, na qual $S(0) = 1$ e $S(+\infty) = \lim_{t \rightarrow \infty} S(t) = 0$. Logo, a função densidade de probabilidade é dada por

$$f(t) = -S'(t).$$

A distribuição de T pode igualmente ser caracterizada pela função de risco:

$$h(t) = \lim_{dt \rightarrow 0^+} \frac{P(t \leq T < t + dt | T \geq t)}{dt}$$

Representa a taxa de morte no instante t , condicional à sobrevivência do indivíduo até esse instante. A função de risco verifica as seguintes propriedades:

$$h(t) \geq 0 \text{ e } \int_0^{\infty} h(t)dt = \infty$$

Sabendo que:

$$P(t \leq T < t + dt | T \geq t) = \frac{P(t \leq T < t + dt)}{P(T \geq t)}$$

Substituindo:

$$\begin{aligned}\lim_{dt \rightarrow 0^+} \frac{\frac{P(t < T \leq t + dt)}{P(T > t)}}{d(t)} &= \frac{1}{P(T > t)} \lim_{dt \rightarrow 0^+} \frac{P(t \leq T < t + dt)}{dt} = \frac{1}{P(T > t)} \times f(t) \\ &= \frac{f(t)}{S(t)} = h(t)\end{aligned}$$

Resumindo, as relações entre a função de sobrevivência, função de densidade de probabilidade e a função de risco são dadas por:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d(\log S(t))}{dt}$$

Como $S(0) = 1$ tem-se que:

$$\begin{aligned}S(t) &= \exp\left(-\int_0^t h(x)dx\right) \\ f(t) &= h(t)\exp\left(-\int_0^t h(x)dx\right)\end{aligned}$$

Por último, a função de risco cumulativa é dada por:

$$H(t) = \int_0^t h(x)dx$$

Logo,

$$H(t) = -\log S(t) \Leftrightarrow S(t) = \exp[-H(t)]$$

$H(t)$ é uma função não negativa e monótona crescente, linear se $h(t)$ for constante, convexa se $h(t)$ crescente e côncava se $h(t)$ decrescente.

Dados censurados referem-se aos indivíduos de que apenas dispomos de informação parcial do seu tempo de vida, visto que durante o período de tempo de estudo não é observado o acontecimento de interesse. Contudo, esta informação, embora parcial, deve ser considerada para possibilitar um estudo mais preciso do acontecimento de interesse. Existem quatro tipos de censura: à direita, à esquerda, intervalar e censura independente ou não informativa:

Censura à direita

Neste tipo de censura apenas se sabe que o tempo de vida excede um determinado valor, uma vez que a observação dos indivíduos termina antes da observação do acontecimento de interesse. Por exemplo, se considerarmos um ensaio

clínico, com data de conclusão pré-definida, cujo evento de interesse é a morte associada a uma determinada doença, os dados censurados à direita irão corresponder aos indivíduos que sobrevivem até à data final, morrem por outra causa (que não tenha a ver com a doença em estudo) ou são perdidos no *follow up*. A censura à direita pode ser tipo I, II ou aleatória. Na censura tipo I, os períodos de observação c_1, \dots, c_n , correspondentes a cada indivíduo são fixados previamente pelo investigador. Neste caso, o número de mortes é aleatório. Se os períodos de observação são iguais, diz-se censura singular, caso contrário trata-se de uma censura múltipla. Na censura tipo II, a observação termina no instante de ocorrência da r -ésima morte, sendo r um número pré-determinado ($1 \leq r \leq n$) e o tempo de estudo é aleatório. Por último, a censura aleatória refere-se aos ensaios clínicos nos quais os indivíduos são incluídos de forma aleatória (por exemplo, de acordo com a data do diagnóstico), e cuja data de conclusão do estudo é pré-definida. Nestes indivíduos observa-se que o tempo decorrido desde que o doente entra no estudo, até ao final deste é aleatório.

Censura à esquerda

Na censura à esquerda apenas é sabido que o tempo de vida é inferior ao tempo de vida registado. Este tipo de censura verifica-se, por exemplo, numa situação que se pretende estudar o tempo até à ocorrência de um tumor em indivíduos expostos a uma substância cancerígena, as observações correspondem aos indivíduos que já apresentam metástases quando são observados são censurados à esquerda. Este tipo de censura é muito menos frequente do que a censura à direita.

Censura intervalar

Este tipo de censura acontece quando não é possível observar o instante exacto no qual ocorre o acontecimento de interesse, apenas é sabido que ocorreu num intervalo aleatório de tempo. Um dos métodos para definir a censura intervalar consiste em averiguar se num determinado instante de monitorização foi observado o acontecimento de interesse. Neste caso os dados designam-se por dados do estado atual. No entanto a censura intervalar, também é utilizada quando apenas se sabe que o acontecimento de interesse ocorre entre dois instantes observados. Este tipo de censura intervalar é comum em estudos longitudinais em que há um *follow up* periódico.

Censura independente ou não informativa

Este tipo de censura pressupõe independência entre os mecanismos de morte e de censura. Neste tipo de censura é exigido que os indivíduos censurados no instante t

sejam representativos de todos os indivíduos, com os mesmos valores nas covariáveis que os indivíduos que sobreviveram até t . É de salientar que, em qualquer instante t , não se pode verificar a censura de indivíduos com base no seu risco de morte.

2.3 Inferência estatística não paramétrica

2.3.1 Estimador da função de sobrevivência

A função de sobrevivência, num dado instante t , pode ser estimada a partir de tempos de vida observados, como sendo a proporção de indivíduos que sobreviveram para além do instante t . Esta função designa-se por função de sobrevivência empírica e é dada por

$$\hat{S}(t) = \frac{\text{número de observações} > t}{n}$$

em que n representa a dimensão da amostra. Contudo, este método não é adequado para estimar a função de sobrevivência a partir de amostras com dados censurados. Assim, (Kaplan e Meier, 1958) propuseram um estimador não paramétrico para a função de sobrevivência quando as amostras incluem dados censurados. Este estimador é designado por estimador de Kaplan-Meier (K-M) ou estimador produto limite. É de salientar que, o estimador de K-M, por ser um método não paramétrico, implica que não existem pressupostos relativos à distribuição da probabilidade do tempo de sobrevivência.

O estimador de K-M é obtido com base na construção de uma sequência de intervalos de tempo, que contemplam um instante de morte no início de cada intervalo. Sejam $t_{(1)}, \dots, t_r$, instantes de morte distintos numa amostra de dimensão n ($r \leq n$), d_i o número de mortes ocorridas em t_i e n_i o número de indivíduos em risco imediatamente antes de t_i . O estimador de K-M da função de sobrevivência é dado por

$$\hat{S}(t) = \prod_{i: t_{(i)} \leq t} \left(\frac{n_i - d_i}{n_i} \right) = \prod_{i: t_{(i)} \leq t} \left(1 - \frac{d_i}{n_i} \right),$$

sendo $\hat{S}(t) = 1$, para $0 \leq t \leq t_{(1)}$. Este estimador trata-se de uma generalização para dados censurados da função de sobrevivência empírica. O estimador de K-M é obtido sob o pressuposto que as mortes são acontecimentos independentes, e que a estimativa da função de sobrevivência em qualquer instante t é a probabilidade de sobreviver para além de t , condicional à probabilidade de ter sobrevivido nos instantes anteriores.

O gráfico da estimativa de K-M da função de sobrevivência é uma função em escada decrescente, na qual a probabilidade de sobrevivência é constante entre os instantes de morte, e diminui a cada morte observada. A estimativa de K-M varia entre 1 e 0. No caso de o maior tempo de vida observado ser $t_{(r)}$, então $\hat{S}(t) = 0$, para $t \geq t_{(r)}$. Se o maior tempo de vida registrado é censurado, t^* , então $\hat{S}(t)$ é indefinida para $t \geq t^*$, pelo que $\hat{S}(t)$ nunca toma o valor zero.

O estimador da função de sobrevivência de K-M é um estimador consistente de $S(t)$, visto aproximar-se da verdadeira função de sobrevivência. Sob condições de regularidade pode ser considerado um estimador de máxima verosimilhança não paramétrico de $S(t)$.

A estimativa da variância de $\hat{S}(t)$ define-se pela fórmula de Greenwood:

$$\widehat{var}\{\hat{S}(t)\} = [\hat{S}(t)]^2 \sum_{i:t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

Logo, o erro padrão da estimativa de K-M é dado por:

$$se\{\hat{S}(t)\} \approx \hat{S}(t) \left\{ \sum_{i:t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)} \right\}^{\frac{1}{2}}$$

Depois de estimado o erro padrão de $\hat{S}(t)$, é possível construir um intervalo de confiança para a função de sobrevivência, num dado instante t . Como temos uma estimativa da função de sobrevivência para cada unidade de tempo, pode-se construir um intervalo de confiança para cada uma dessas unidades de tempo. Este intervalo é obtido assumindo que o valor da estimativa da função de sobrevivência num determinado instante t , apresenta uma distribuição normal de valor médio $\hat{S}(t)$ e $var\{\hat{S}(t)\}$. Assim sendo, um intervalo de 100 $(1-\alpha)\%$ de confiança para $S(t)$ é dado por

$$\left[\hat{S}(t) - z_{\alpha/2} \sqrt{var\{\hat{S}(t)\}}, \hat{S}(t) + z_{\alpha/2} \sqrt{var\{\hat{S}(t)\}} \right],$$

em que $z_{\alpha/2}$ representa o quantil de probabilidade $1 - \frac{\alpha}{2}$ da distribuição normal centrada e reduzida, ou seja, da distribuição $N(0,1)$.

No caso do estimador de K-M, o intervalo de 100 $(1-\alpha)\%$ de confiança para $S(t)$, para um dado instante t , é dado por:

$$[\hat{S}(t) - z_{\alpha/2} se\{\hat{S}(t)\}, \hat{S}(t) + z_{\alpha/2} se\{\hat{S}(t)\}].$$

Contudo, através deste método é obtido um intervalo de confiança simétrico, que não é considerado apropriado quando a estimativa da função de sobrevivência

aproxima-se de zero ou da unidade, visto que podem ser obtidos limites de confiança da função de sobrevivência fora do intervalo (0,1). Para resolver esta discrepância pode ser considerada a substituição de qualquer limite superior a 1, por 1, e de qualquer limite inferior a 0 por 0. Em alternativa, pode proceder-se à transformação de $\hat{S}(t)$, com a finalidade de obter um valor no domínio $(-\infty, +\infty)$ e obter um intervalo de confiança para o valor transformado. As transformações possíveis incluem a transformação logística, $\log\left(\frac{S(t)}{1-S(t)}\right)$ e a transformação complementar, $\log\{-\log S(t)\}$. A transformação complementar permite assegurar que os limites de confiança são positivos e menores ou iguais a 1 e trata-se de um método preciso porque $\log\{-\log S(t)\}$ apresenta uma distribuição mais aproximada da distribuição normal do que $\hat{S}(t)$. Então, o intervalo de 100 $(1-\alpha)\%$ confiança para $\log\{-\log S(t)\}$ define-se por:

$$\left[W - z_{\alpha/2} \sqrt{\text{var}(W)}, W + z_{\alpha/2} \sqrt{\text{var}(W)} \right],$$

em que $W = \log\{-\log S(t)\}$, $z_{\alpha/2}$ representa o quantil de probabilidade $1 - \frac{\alpha}{2}$ da distribuição $N(0,1)$ e $\text{var}(W) = \frac{1}{\{\log S(t)\}^2} \sum_{i: t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}$.

Dado que, geralmente a distribuição do tempo de vida é assimétrica positiva, é preferível utilizar a mediana para caracterizar a localização da distribuição. Então sendo $\hat{S}(t)$ a estimativa de K-M para a função de sobrevivência, a estimativa da mediana do tempo de vida é definida por:

$$m = \min\{t_i: \hat{S}(t_i) \leq 0.5\}$$

onde t_i é o i -ésimo instante de morte, $i = 1, \dots, k$.

Em doenças com prognóstico favorável, acontece por vezes que a estimativa da função de sobrevivência é superior a 0.5, para todos os valores de t . Nesse caso, não é possível obter uma estimativa não paramétrica da mediana do tempo de vida.

2.3.2 Estimador da função de risco cumulativa

Quando se pretende estimar a função de risco cumulativa, $H(t)$, um estimador natural é $\hat{H}(t) = -\log \hat{S}(t)$, onde $\hat{S}(t)$ é o estimador de K-M. Contudo, uma alternativa ao método anterior é o estimador de Nelson-Aalen (N-A). Sejam $t_{(1)}, \dots, t_{(r)}$, instantes de morte distintos numa amostra de dimensão n ($r \leq n$), d_i o número de mortes ocorridas em t_i e n_i o número de indivíduos em risco no instante anterior a t_i . O estimador de N-A, define-se da seguinte forma:

$$\tilde{H}(t) = \sum_{i:t_{(i)} \leq t} \frac{d_i}{n_i}$$

O estimador de N-A, para além de permitir obter uma estimativa da função de risco cumulativa, possibilita igualmente estimar a função de sobrevivência, visto que $\hat{H}(t) = -\log \hat{S}(t)$, logo $\tilde{S}(t) = \exp(-\tilde{H}(t))$. Neste caso o estimador de N-A, para a função de sobrevivência, também conhecido por estimador de Breslow é dado por:

$$\hat{S}_{NA}(t) = \exp\left(-\sum_{i:t_{(i)} \leq t} \frac{d_i}{n_i}\right)$$

Sabendo que:

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \dots,$$

Tem-se que e^{-x} , é aproximadamente igual a $1 - x$, quando x é pequeno. Considerando que $x = d_i/n_i$, logo $\exp(-d_i/n_i) \approx 1 - d_i/n_i = (n_i - d_i)/n_i$, desde que d_i seja pequeno comparativamente a n_i (que normalmente será, com a exceção dos últimos instantes de morte). Nestas condições o estimador de K-M aproxima-se do estimador da função de sobrevivência de N-A. Em amostras pequenas o estimador da função de sobrevivência de N-A parece apresentar um comportamento melhor, no entanto na maioria das circunstâncias será obtido um resultado semelhante com o estimador da função de sobrevivência de K-M. Contudo, como o estimador de K-M é uma generalização da função de sobrevivência empírica a sua utilização será preferível.

2.3.3 Testes para comparação de curvas de sobrevivência

Quando se pretende comparar a função de sobrevivência de dois ou mais grupos, a representação gráfica da estimativa de K-M estratificada pode ser útil para analisarmos o comportamento das diferentes curvas de sobrevivência. Contudo, para verificar se existe uma diferença estatisticamente significativa entre várias curvas de sobrevivência é necessário o recurso ao teste de hipóteses. Se considerarmos duas amostras com m e n indivíduos, provenientes de duas populações com função de sobrevivência $S_1(t)$ e $S_2(t)$, respetivamente, pretendemos testar as seguintes hipóteses:

$$H_0: S_1(t) = S_2(t) \quad vs. \quad H_1: S_1(t) \neq S_2(t)$$

Existem diversos testes não paramétricos adequados para proceder a esta comparação, sendo de destacar o teste log-rank e de Gehan.

Teste log-rank

O teste log-rank possibilita a análise do desvio entre o número de mortes observadas e o número de mortes esperadas sob H_0 . Sejam $t_1 < \dots < t_k$ instantes de morte distintos relativos aos $m + n$ indivíduos; d_j o número de mortes em $t_j, j = 1, \dots, k$; d_{ij} o número de mortes em t_j no grupo $i, i = 1, 2$; n_j o número de indivíduos em risco em $t_j, j = 1, \dots, k$; $n_{i,j}$ o número de indivíduos em risco em t_j , no grupo $i, i = 1, 2$. A informação relevante em cada instante t_j pode ser resumida na seguinte tabela de contingência:

Grupo	Número de mortes	Número de indivíduos que sobrevive para além de t_j	Número de indivíduos em risco em t_j
I	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
II	d_{2j}	$n_{2j} - d_{2j}$	n_{2j}
Total	d_j	$n_j - d_j$	n_j

Mantel e Haenszel (1959) consideraram a distribuição das frequências observadas de cada célula, dados os totais marginais, sob a validade da hipótese nula. Então, supondo que H_0 é verdadeira, a distribuição de d_{1j} , condicional aos valores marginais é hipergeométrica e é dada por:

$$p(d_{1j}|d_j, n_j) = \frac{\binom{d_j}{d_{1j}} \binom{n_j - d_j}{n_{1j} - d_{1j}}}{\binom{n_j}{n_{1j}}}$$

O valor médio e variância condicionais a d_{1j} são respetivamente:

$$e_{1j} = \frac{n_{1j}d_j}{n_j} \quad v_{1j} = \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)},$$

sendo e_{1j} o número esperado de mortes no instante t_j no grupo 1. Sob a hipótese nula, a probabilidade de morte no instante t_j não depende do grupo no qual o indivíduo se encontra, logo a probabilidade de morte no instante t_j é d_j/n_j . Multiplicando esta probabilidade pelo número de indivíduos no grupo I, obtemos o valor esperado para o grupo 1.

O passo seguinte passa por combinar a informação de todas as tabelas de contingência, de forma a obter uma medida global do desvio entre os valores observados e os valores esperados de d_{1j} . A forma mais direta de obter esta medida será fazer a diferença entre d_{1j} e e_{1j} , para cada instante de morte. A estatística resultante e a respetiva variância é dada por:

$$U_L = \sum_{j=1}^r (d_{1j} - e_{1j})$$

$$V_L = Var(U_L) = \sum_{j=1}^r v_{1j}$$

A estatística U_L segue uma distribuição normal, quando o número de mortes não é muito pequeno e $U_L/\sqrt{V_L}$ apresenta uma distribuição $N(0,1)$.

Por último, a estatística $W_L = U_L^2/V_L$ reflecte a dimensão do desvio entre o número de mortes observadas e esperadas sob H_0 . Quanto maior a estatística de teste, maior a evidência para rejeitar H_0 . Atendendo a que a estatística de teste tem uma distribuição aproximada de qui-quadrado com um grau de liberdade, o valor-p associado pode ser obtido a partir da distribuição de qui-quadrado.

Teste de Gehan

O teste de Gehan trata-se de uma generalização do teste de Mann-Whitney-Wilcoxon para dados censurados. Este teste também é conhecido por teste de Wilcoxon generalizado e é utilizado para testar a hipótese nula, que não existe diferença nas funções de sobrevivência de dois grupos.

Sejam m e n as dimensões das amostras correspondentes aos grupos 1 e 2 respetivamente. Considere-se uma amostra conjunta com $m+n$ tempos de observação ordenados por ordem crescente e δ_i uma variável indicatriz. Seja $U_{kj} = U(t_k, t_j)$ a pontuação atribuída ao comparar um tempo fixo t_k , com os tempos observados. Então:

$$U_{kj} = \begin{cases} +1 & \text{se } (t_k > t_j, \delta_j = 1) \text{ ou } (t_k = t_j, \delta_k = 0, \delta_j = 1) \\ -1 & \text{se } (t_k < t_j, \delta_k = 1) \text{ ou } (t_k = t_j, \delta_k = 1, \delta_j = 0) \\ 0 & \text{caso contrário} \end{cases}$$

Seja $U'_k = \sum_{j=1}^{m+n} U_{kj}$ para $k = 1, \dots, m+n$ ($j \neq k$).

Portanto, para cada observação, $k = 1, \dots, m + n$ é atribuída uma pontuação U'_k , que é igual à diferença entre o número das restantes $m + n - 1$ e o número de observações que correspondem a tempos de vida que são de certeza maiores que t_k .

A estatística resultante é dada por

$$U = \sum_{k=1}^{m+n} U'_k \text{ para } k: t_k \in \text{amostra } 1,$$

sob a validade de H_0

$$E(U) = 0 \text{ e } var = \frac{mn}{(m+n)(m+n-1)} \sum_{k=1}^{m+n} (U'_k)^2.$$

Podemos então considerar a estatística $Z = U / \sqrt{var((U))}$, que sob H_0 , tem distribuição assintótica $N(0,1)$. Uma forma equivalente para a estatística de teste é dada por

$$U_G = \sum_{j=1}^r n_j (d_{1j} - e_{1j})$$

onde $e_{1j} = n_{1j}d_j/nj$. A variância da estatística U_G é dada por

$$V_G = \sum_{j=1}^r n_j^2 v_{1j}$$

e a estatística de teste de Gehan é então

$$W_G = U_G^2 / V_G.$$

sob a validade de H_0 , W_G tem uma distribuição assintótica de qui-quadrado com 1 grau de liberdade.

Teste log-rank vs. teste de Gehan

O teste log-rank é o mais potente na detecção de afastamentos da hipótese de igualdade de distribuições, que sejam do tipo de riscos proporcionais. Em contrapartida, quando existe evidência de que as funções de risco cruzam deve optar-se pelo teste de Gehan, uma vez que o teste log-rank pode não permitir detetar diferenças significativas entre as curvas de sobrevivência. Assim sendo, é importante avaliar a validade da hipótese de riscos proporcionais. Sabe-se que se as funções de risco são proporcionais, então as funções de sobrevivência não se cruzam. Logo, o cruzamento das funções de sobrevivência invalida a hipótese de riscos proporcionais.

Seja $h_1(t)$ a função de risco de um indivíduo no grupo 1 no instante t , e $h_2(t)$ a função de risco de um indivíduo no grupo 2 no mesmo instante. No caso das funções de risco serem proporcionais, $h_1(t) = \varphi h_2(t)$ em que φ não depende de t , tem-se que

$$h_1(t) = \varphi h_2(t) \Rightarrow \exp\left\{-\int_0^t h_1(u)du\right\} = \exp\left\{-\int_0^t \varphi h_2(u)du\right\},$$

como,

$$S(t) = \exp\left(-\int_0^t h(u)du\right)$$

então,

$$S_1(t) = [S_2(t)]^\varphi.$$

Assim sendo, conclui-se com base neste resultado que $S_1(t)$ é maior ou menor que $S_2(t)$ consoante φ , seja maior ou menor do que a unidade, para qualquer instante t . Isto significa que se as duas funções de risco são proporcionais, as verdadeiras funções de sobrevivência não se cruzam.

Uma avaliação da hipótese de riscos proporcionais pode ser efetuada, com base na representação gráfica das estimativas da função de sobrevivência. Se as estimativas das funções de sobrevivência não se cruzam, considera-se que a proporcionalidade de riscos existe e como tal é adequado usar o teste log-rank. Outro método gráfico, mais rigoroso para averiguar a proporcionalidade das funções de risco, é a representação gráfica das funções $\log[-\log \hat{S}(t)]$, para cada grupo. Se, $h_1(t)$ e $h_2(t)$ são funções de risco proporcionais então:

$$\begin{aligned} S_1(t) &= [S_2(t)]^\varphi \Rightarrow -\log S_1(t) = \varphi[-\log S_2(t)] \Rightarrow \\ &\Rightarrow \log[-\log S_1(t)] = \log \varphi + \log[-\log S_2(t)]. \end{aligned}$$

Deste modo, o logaritmo das funções de risco cumulativas de dois indivíduos pertencentes a dois grupos diferentes, apresentam uma distância constante e igual a $\log \varphi$. Então, se $\hat{S}_1(t)$ e $\hat{S}_2(t)$ são estimativas de $S_1(t)$ e $S_2(t)$, o gráfico $\log[-\log \hat{S}_1(t)]$ versus t tenderá a ser paralelo ao gráfico de $\log[-\log \hat{S}_2(t)]$ versus t , quando $h_1(t)$ e $h_2(t)$ são proporcionais. Assim sendo, deverá ser avaliado se a distância entre os gráficos se mantém razoavelmente constante, ao longo do tempo.

Classe de testes não paramétricos

Esta classe engloba os testes não paramétricos para os quais a estatística de teste é dada por:

$$\frac{[\sum_{j=1}^r w_j (d_{1j} - e_{1j})]^2}{\sum_{j=1}^r w_j^2 v_{1j}}$$

onde w_j são constantes conhecidas. Sob H_0 , esta estatística tem uma distribuição assintótica de qui-quadrado com um grau de liberdade. De acordo com os valores atribuídos aos pesos w_j , obtemos diferentes testes, nomeadamente:

$w_j = 1$ teste log-rank

$w_j = n_j$ teste de Gehan

$w_j = \sqrt{n_j}$ teste de Tarone-Ware

$w_j = \prod_{i:t(i) \leq t(j)} \left(1 - \frac{d_i}{n_{i+1}}\right)$ teste de Peto-Peto

O teste de Tarone-Ware é um compromisso entre o teste de Gehan e o teste log-rank, visto que atribui também maior peso às diferenças na fase inicial, embora em menor proporção do que o teste de Gehan. No caso do teste de Peto-Peto, o peso é um estimador da função de sobrevivência comum aos dois grupos.

Comparação de três ou mais grupos

Tanto o teste de Gehan como o teste log-rank são extensíveis a três ou mais grupos. Na análise de 3 ou mais grupos, são definidos análogos das estatísticas U, para comparação do número observado de mortes nos grupos $1, 2, \dots, g - 1$ com o número de mortes esperadas.

$$U_{Lk} = \sum_{j=1}^r \left(d_{jk} - \frac{n_{kj} d_j}{n_j} \right)$$

$$U_{Gk} = \sum_{j=1}^r n_j \left(d_{jk} - \frac{n_{kj}d_j}{n_j} \right)$$

Seja g o número de grupos a comparar e $k = 1, 2, \dots, g - 1$. As quantidades U_{Lk} e U_{Gk} são expressas em forma de vetor com $g - 1$ componentes, e referem-se às estatísticas de teste obtidas pelo teste log-rank e Gehan, respetivamente. As expressões para a variância e covariância entre pares para o teste log-rank é definida da seguinte forma:

$$V_{Lkk'} = \sum_{j=1}^r \frac{n_{kj}d_j(n_j - d_j)}{n_j(n_j - 1)} \left(\delta_{kk'} - \frac{n_{k'j}}{n_j} \right)$$

para $k, k' = 1, 2, \dots, g - 1$, onde $\delta_{kk'} = \begin{cases} 1 & \text{se } k = k' \\ 0 & \text{Caso contrário} \end{cases}$

Os valores obtidos são agrupados numa matriz de covariância V_L . De forma semelhante a matriz de covariância do teste de Gehan é a matriz V_G , cujo elemento kk' é dado por:

$$V_{Gkk'} = \sum_{j=1}^r n_j \frac{n_{kj}d_j(n_j - d_j)}{n_j(n_j - 1)} \left(\delta_{kk'} - \frac{n_{k'j}}{n_j} \right).$$

Finalmente, o teste da hipótese nula é feito com base no resultado das estatísticas:

$$U_L^{-1} V_L^{-1} U_L \text{ ou } U_G^{-1} V_G^{-1} U_G.$$

Sob H_0 , qualquer uma das estatísticas apresenta distribuição qui-quadrado com $g - 1$ graus de liberdade.

2.4 Modelo de Cox

Num modelo de riscos proporcionais onde $\varphi(x_i)$ é uma função do vetor das variáveis explicativas para o i -ésimo indivíduo e $h_0(t)$ expressa a função de risco subjacente, a função de risco no instante t é dada por:

$$h_i(t) = \varphi(x_i) h_0(t)$$

A função $\varphi(\cdot)$ pode ser interpretada como o risco no instante t para um indivíduo cujo vetor de variáveis explicativas é x_i , em relação a um indivíduo para o qual $x = 0$. Atendendo a que $\varphi(x_i)$ não pode ser negativo, é conveniente considerar $\exp(\eta_i)$, onde η_i é uma combinação linear das variáveis explicativas. Assim,

$$\eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}$$

A quantidade η_i é a componente linear do modelo, também conhecida por índice de prognóstico para o i -ésimo indivíduo. Assim, o modelo de riscos proporcionais é dado por:

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}) h_0(t)$$

em que β_1, \dots, β_p são os coeficientes de regressão (desconhecidos) que representam o efeito das covariáveis no tempo de vida. Após reformulação, o modelo pode ser apresentado da seguinte forma

$$\log \left\{ \frac{h_i(t)}{h_0(t)} \right\} = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi},$$

donde o modelo de riscos proporcionais pode ser tido como um modelo linear do logaritmo do risco relativo. A função de risco subjacente $h_0(t)$ representa a função de risco para o indivíduo a que está associado o vetor $x = 0$. No modelo proposto por (Cox, 1972) não são considerados pressupostos quanto à distribuição da função de risco subjacente, os coeficientes são estimados sem essa informação.

2.4.1 Função de verosimilhança parcial

Considere-se que se encontram em estudo n indivíduos e que foram observados r tempos de vida distintos $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$, tal que t_j é o j -ésimo tempo de vida ordenado e há $n - r$ tempos de censura à direita. Assumindo que não existem dados empatados nas observações, seja $R(t) = R(t_{(j)}) = \{j: t_j \geq t_{(i)}\}$ o conjunto de indivíduos em risco no instante $t_{(j)}$ e $x_{(j)}$ o vetor de covariáveis associado ao indivíduo que morreu no instante $t_{(j)}$.

Cox baseou a inferência sobre β , na seguinte função:

$$L(\beta) = \prod_{j=1}^r \frac{e^{(\beta' x_{(j)})}}{\sum_{l \in R(t_{(j)})} e^{(\beta' x_l)}}$$

É de notar que a função de verosimilhança depende apenas dos *rankings* dos tempos de vida, visto que é esta informação que determina o número de indivíduos em risco em cada instante. Consequentemente, a inferência sobre o efeito das covariáveis depende apenas das ordens dos tempos de vida.

É importante salientar que esta função de verosimilhança não depende de $h_0(t)$ e permite portanto a inferência sobre o vetor dos parâmetros β . Esta função não é uma verosimilhança no sentido usual, contudo pode ser interpretada como uma

verossimilhança parcial, visto que permite a realização de inferência na presença de parâmetros perturbadores, que neste caso se referem a $h_0(t)$.

O estimador de máxima verossimilhança parcial de β é consistente e assintoticamente normal com valor médio β e matriz de covariância $I(\beta)^{-1}$, onde

$$I_{jk} = -E \left(\frac{\delta^2 \log L}{\delta \beta_j \delta \beta_k} \right).$$

2.4.2 Existência de observações empatadas

Quando ocorre um tempo observado e um tempo censurado em simultâneo admite-se que o tempo censurado, ocorre depois do tempo observado. Por vezes podem ocorrer dois tempos observados ou censurados no mesmo instante, neste caso, é necessário alterar a função de verossimilhança para acomodar empates.

Considere-se n indivíduos para os quais foram observados, tempos de vida distintos $t_{(1)}, \dots, t_{(k)}$. Seja x_{ij} o vetor de covariáveis associado ao indivíduo j , $j = 1, \dots, d_i$ que morre em $t_{(i)}$. Se o número d_i de indivíduos que morrem em $t_{(i)}$ é pequeno comparado com o número de indivíduos pertencentes a R_i , pode ser utilizada a aproximação da função de verossimilhança proposta por Peto (1972) e Breslow (1974):

$$L(\beta) = \prod_{i=1}^k \frac{e^{(\beta' s_{(i)})}}{\left\{ \sum_{l \in R(t_{(i)})} e^{(\beta' x_l)} \right\}^{d_i}}$$

onde $s_i = \sum_{j=1}^{d_i} x_{ij}$ para $i = 1, \dots, k$.

2.4.3 Intervalos de confiança e testes de hipóteses para β

O intervalo de confiança a $100(1-\alpha)\%$ para um parâmetro β , é um intervalo com limites

$$\hat{\beta} \pm z_{\alpha/2} se(\hat{\beta}),$$

onde $\hat{\beta}$ é uma estimativa de β , e $z_{\alpha/2}$ é o quantil $1 - \alpha/2$ da distribuição $N(0,1)$. As hipóteses $H_0: \beta = 0$ vs. $H_1: \beta \neq 0$ podem ser testadas com base na estatística

$$\hat{\beta}^2 / var(\hat{\beta}),$$

que corresponde ao teste de Wald. Na interpretação do valor-p, para um dado parâmetro β_j é importante considerar que a hipótese que está a ser testada é de que

$\beta_j = 0$, na presença dos restantes termos no modelo. As estimativas individuais dos β 's no modelo de riscos proporcionais não são independentes, por isso são de difícil interpretação. Assim sendo, existem metodologias mais adequadas para o teste de hipóteses do que o teste de Wald.

O risco relativo é dado por $\psi = e^{\beta}$ sendo a estimativa do risco relativo dado por $\hat{\psi} = e^{\hat{\beta}}$ e o erro padrão de $\hat{\psi}$, pode ser obtido a partir do erro padrão de $\hat{\beta}$. Assim, a variância de $\hat{\psi}$ é uma função de $\hat{\beta}$ dada por

$$\{e^{\hat{\beta}}\}^2 \text{var}(\hat{\beta})$$

que corresponde a $\hat{\psi}^2 \text{var}(\hat{\beta})$, logo o erro padrão de $\hat{\psi}$ é dado por:

$$se(\hat{\psi}) = \hat{\psi} se(\hat{\beta}).$$

2.4.4 Estimação da função de sobrevivência

Dado que

$$S(t; x) = [S_0(t)]^{\exp(\beta'x)}$$

ao estimarmos $S_0(t)$ torna-se possível obter estimativas de $S(t; x)$ para qualquer x . Tendo obtido $\hat{\beta}$ a partir da verosimilhança parcial, a função de sobrevivência $S_0(t)$ pode ser determinada com base num estimador de máxima verosimilhança proposto por Kalbfleisch e Prentice.

Considere-se n indivíduos nos quais foram observados k tempos de vida distintos $t_{(1)} < \dots < t_{(k)}$, $k < n$; R_i o conjunto de risco no instante $t_{(i)}$ e $D_{(i)}$ o conjunto de índices associados a d_i indivíduos que morreram em $t_{(i)}$. Atendendo a que a função de risco em $t_{(i)}$, $i = 1, \dots, k$ é $h_i = 1 - \alpha_i$ com $\alpha_i = S_0(t_{(i+1)})/S_0(t_{(i)})$, considerando $\beta = \hat{\beta}$, são obtidas as seguintes equações de máxima verosimilhança:

$$\sum_{l \in D_i} \frac{\exp(\hat{\beta}'x_l)}{1 - \alpha_i} = \sum_{l \in R_i} \exp(\hat{\beta}'x_l),$$

quando $d_i = 1$, $i = 1, \dots, k$ obtém-se:

$$\hat{\alpha} = \left(1 - \frac{\exp(\hat{\beta}'x_{(i)})}{\sum_{l \in R_i} \exp(\hat{\beta}'x_l)} \right)^{\exp(-\hat{\beta}'x_{(i)})}$$

Caso contrário, é necessário recorrer a um método iterativo. O estimador de máxima verosimilhança de $S_0(t)$ é dado por:

$$\hat{S}_0(t) = \prod_{i:t_{(i)} \leq t} \hat{a}_i$$

Breslow propôs um estimador que não requer a utilização de métodos iterativos quando $d_i > 1$, para algum i :

$$\begin{aligned} \tilde{H}_0(t) &= -\log \tilde{S}_0(t) \\ &= \sum_{i:t_{(i)} \leq t} \frac{d_i}{\sum_{l \in R_i} \exp(\hat{\beta}' x_l)} \end{aligned}$$

$\tilde{S}_0(t)$ e $\hat{S}_0(t)$ não diferem muito quando $d_i = 1, i = 1, \dots, k$ ou quando existem poucas observações exceto na cauda direita da distribuição.

2.4.5 Comparação de distribuições do tempo de vida

O modelo de Cox permite testar a hipótese de igualdade nas distribuições do tempo de vida de dois grupos de indivíduos, contra a hipótese alternativa de que as distribuições são diferentes, desde que apresentem funções de risco proporcionais.

Seja x uma covariável binária que toma o valor zero se o indivíduo pertence ao grupo 1, e a unidade, caso pertença ao grupo 2. As funções de sobrevivência correspondentes aos dois grupos estão relacionadas por:

$$S_2(t) = S_1(t)^{\exp(\beta)}$$

De notar que testar $H_0: S_1(t) = S_2(t)$ é equivalente a testar $H_0: \beta = 0$ vs. $H_1: \beta \neq 0$.

Sejam $t_1 < \dots < t_k$ tempos de vida distintos relativos aos $m + n$ pacientes; d_j o número de mortes ocorridas em $t_j, j = 1, \dots, k$; d_{ij} o número de mortes ocorridas em t_j no grupo $i, i = 1, 2$; n_j o número de indivíduos em risco em $t_j, j = 1, \dots, k$; n_{ij} o número de indivíduos em risco em t_j no grupo $i, i = 1, 2$. Sob o modelo de Cox e supondo que existem poucas observações empatadas tem-se que

$$\log L(\beta) = r_2 \beta - \sum_{j=1}^k d_j \log(n_{1j} + n_{2j} e^{\beta})$$

onde $r_2 = \sum_{j=1}^k d_{2j}$,

como,

$$U(\beta) = \frac{\partial \log L}{\partial \beta} = r_2 - \sum_{j=1}^k \frac{d_j n_{2j} e^\beta}{n_{1j} + n_{2j} e^\beta}$$

$$I(\beta) = \frac{\partial^2 \log L}{\partial \beta^2} = \sum_{j=1}^k \frac{d_j n_{2j} e^\beta}{(n_{1j} + n_{2j} e^\beta)^2}$$

Sob $H_0: \beta = 0$, a estatística $Z = \frac{U(0)}{\sqrt{I(0)}}$ tem distribuição assintótica $N(0,1)$. Sabendo que:

$$U(0) = \sum_{j=1}^k (d_{2j} - \frac{d_j n_{2j}}{n_j}) \quad I(0) = \sum_{j=1}^k \frac{d_j n_{1j} n_{2j}}{n_j^2}$$

Quando existem muitas observações empatadas, deve ser utilizado um teste que acomode a natureza discreta dos dados. Este teste é também baseado na estatística Z , com $U(0)$ dado como anteriormente, mas com $I(0)$ dado por:

$$I(0) = \sum_{j=1}^k \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$$

Sob H_0 , Z^2 tem distribuição assintótica de qui-quadrado com um grau de liberdade. Este teste é equivalente ao teste de log-rank e por vezes é designado por teste de Cox-Mantel.

2.4.6 Método para a seleção de covariáveis

Estratégia recomendada para a seleção de covariáveis por (Collett, 2003):

1. O primeiro passo consiste na construção de modelos, nos quais é introduzida uma covariável de cada vez. Posteriormente, os valores da estatística $-2 \log \hat{L}$ são comparadas com o modelo nulo para determinar que covariáveis em isolado reduzem significativamente a estatística $-2 \log \hat{L}$.
2. As covariáveis que parecem importantes no passo 1 são incluídas no mesmo modelo. As covariáveis que não aumentam significativamente o valor de $-2 \log \hat{L}$ quando omitidas, podem ser descartadas. Apenas as covariáveis que levam a um aumento significativo da estatística de teste é que devem ser mantidas no modelo.
3. As covariáveis que não foram consideradas importantes em isolado, e que não entraram no passo 2, são testadas na presença de outras covariáveis. Estas covariáveis são assim adicionadas ao modelo obtido no passo 2, uma de cada

vez, se alguma reduzir significativamente o valor da estatística $-2 \log \hat{L}$, deverá ser incluída no modelo.

4. Uma análise final é realizada para assegurar que nenhuma covariável possa ser omitida sem produzir um aumento significativo do valor da estatística $-2 \log \hat{L}$, e que nenhuma covariável não incluída reduza significativamente o valor de $-2 \log \hat{L}$.

É recomendado um nível de significância de 10%, na decisão de omitir ou adicionar uma covariável ao modelo.

Contudo, a estatística $-2 \log \hat{L}$, não permite por si só avaliar a adequabilidade do modelo, visto que o valor de \hat{L} depende da dimensão da amostra. Assim, $-2 \log \hat{L}$ apenas permite a comparação de modelos ajustados aos mesmos dados.

Na comparação de modelos que não estão aninhados, pode ser utilizada a seguinte estatística

$$AIC = -2 \log \hat{L} + \alpha q,$$

onde q representa o número de parâmetros β desconhecidos no modelo e α é uma constante pré-determinada. Esta estatística é conhecida por Critério de Informação de Akaike, quanto menor for o valor desta estatística, melhor será o ajustamento do modelo aos dados.

A constante α , normalmente toma um valor entre 2 e 6. A escolha de $\alpha = 3$ corresponde aproximadamente a usar-se uma significância de 5%. De um modo geral este valor é o mais recomendado.

2.4.7 Comparação de modelos alternativos

Seja o modelo (1), composto por um subconjunto dos termos do modelo (2), assim diz-se que o modelo (1) está aninhado no modelo (2). Especificamente, se considerarmos p covariáveis explicativas, x_1, x_2, \dots, x_p que foram incluídas no modelo (1), a função de risco será dada por:

$$h(t) = \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) h_0(t)$$

Considere-se também que $p + q$ variáveis explicativas $x_1, x_2, \dots, x_p, x_{p+1}, \dots, x_{p+q}$ estão incluídas num modelo (2), sendo a expressão obtida:

$$h(t) = \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \beta_{p+1} x_{p+1} + \dots + \beta_{p+q} x_{p+q}) h_0(t)$$

O modelo (2) contém assim q covariáveis adicionais $x_{p+1}, x_{p+2}, \dots, x_{p+q}$. Atendendo a que o modelo (2) apresenta um maior número de covariáveis do que o modelo (1), o modelo (2) deve ter um melhor ajustamento aos dados. No entanto deve ser determinado até que ponto as q covariáveis adicionais melhoram o poder explicativo do modelo. Caso possam ser omitidas, o modelo (1) será mais adequado.

O efeito de um determinado termo incluído no modelo depende dos restantes termos nesse modelo. Por exemplo, no modelo (1) o efeito de qualquer uma das p covariáveis explicativas incluídas na função de risco depende das $p - 1$ covariáveis incluídas no modelo, assim o efeito de x_p , diz-se ser ajustado para as restantes $p - 1$ covariáveis. De uma forma semelhante quando acrescentamos as q covariáveis adicionais $x_{p+1}, x_{p+2}, \dots, x_{p+q}$, no modelo (2), o efeito destas covariáveis na função de risco é ajustado para as p covariáveis já incluídas no modelo x_1, x_2, \dots, x_p .

Sejam $\hat{L}(1)$ e $\hat{L}(2)$ verosimilhanças maximizadas para cada modelo, os dois modelos podem ser comparados com base na diferença entre os valores de $-2 \log \hat{L}$ para cada modelo. A existência de uma grande diferença entre $-2 \log \hat{L}(1)$ e $-2 \log \hat{L}(2)$ levará à conclusão que as q covariáveis adicionais no modelo (2), melhoram a adequação do modelo. A alteração da estatística $-2 \log \hat{L}$ em função da inclusão de novos termos irá depender dos termos que já estavam incluídos no modelo.

A diferença entre $-2 \log \hat{L}(1)$ e $-2 \log \hat{L}(2)$ é dada por $-2 \log \hat{L}(1) + 2 \log \hat{L}(2)$ que por sua vez expressa o efeito de acrescentar as covariáveis $x_{p+1}, x_{p+2}, \dots, x_{p+q}$ a um modelo que já contém x_1, x_2, \dots, x_p . Por outras palavras, diz-se ser a alteração do valor $-2 \log \hat{L}$ devido à inclusão dos termos $x_{p+1}, x_{p+2}, \dots, x_{p+q}$, ajustado para x_1, x_2, \dots, x_p .

A estatística $-2 \log \hat{L}(1) + 2 \log \hat{L}(2)$ pode também ser dada por:

$$-2 \log \left\{ \frac{\hat{L}(1)}{\hat{L}(2)} \right\}$$

que se trata da estatística do teste da razão de verosimilhanças para testar a hipótese nula de que os q parâmetros $\beta_{p+1}, \beta_{p+2}, \dots, \beta_{p+q}$, no modelo (2), são todos iguais a zero. Esta estatística tem uma distribuição assintótica de qui-quadrado, sob a hipótese nula de que todos os β adicionais são zero. O número de graus de liberdade é igual à diferença entre o número parâmetros β a serem testados nos dois modelos. Assim sendo, na comparação do modelo (1) e modelo(2), utiliza-se estatística $-2 \log \hat{L}(1) + 2 \log \hat{L}(2)$, que apresenta uma distribuição de qui-quadrado com q graus de liberdade, sob a hipótese nula de que $\beta_{p+1}, \beta_{p+2}, \dots, \beta_{p+q}$ são todos zero. Se o valor da estatística não for suficientemente elevado, considera-se que ambos os modelos são adequados. Neste caso, o modelo com menor número de termos deve ser

considerado melhor. No caso oposto, em que existe um valor de estatística de teste muito elevado, considera-se que os termos adicionais são importantes para o modelo, assim adoptar-se-ia o modelo mais complexo.

2.4.8 Métodos de diagnóstico para o modelo de Cox

Resíduos de Cox-Snell

O resíduo de Cox-Snell é definido para o i -ésimo indivíduo como

$$r_i = \hat{H}(t_i) = \exp(\hat{\beta}x_i)\hat{H}_0(t_i),$$

sendo $\hat{\beta}$ e \hat{H}_0 as estimativas de máxima verosimilhança parcial. Este resíduo é útil para avaliar o ajustamento global do modelo, visto que se for obtido um bom ajustamento então os valores estimados $\hat{H}(t_i)$ terão propriedades semelhantes aos verdadeiros valores $H(t_i)$. Isto é, deverão comportar-se como uma amostra aleatória proveniente de uma população com distribuição $\exp(1)$. No entanto, atendendo a que é necessário considerar a existência de observações não observadas, então os resíduos de Cox-Snell deverão comportar-se com uma amostra censurada de uma distribuição $\exp(1)$. Assim sendo, é de ter em conta os resíduos de Cox Snell modificados, dados por:

$$r'_i = \begin{cases} r_i & \text{se } t_i \text{ é uma observação não censurada} \\ r_i + 1 & \text{se } t_i \text{ é uma observação censurada} \end{cases}$$

Assim, no caso de o modelo ser adequado, no gráfico dos resíduos de Cox-Snell modificados *versus* a estimativa de Nelson-Aalen para a função de risco cumulativa é obtida uma linha reta com declive 1 e ordenada na origem nula.

Resíduos de Schoenfeld

O resíduo de Schoenfeld para o i -ésimo indivíduo correspondente à covariável x_j , onde $j = 1, \dots, p$, é dado por

$$r_{ji} = \delta_i \{x_{ji} - a_{ij}\}$$

onde,

$$a_{ij} = \frac{\sum_{l \in R_i} x_{jl} \exp(\hat{\beta}' x_l)}{\sum_{l \in R_i} \exp(\hat{\beta}' x_l)}$$

Assim, o resíduo é a diferença entre o valor da covariável x_j para um indivíduo para o qual se verificou o acontecimento de interesse num determinado instante t_i , e uma média ponderada dos valores para essa covariável para todos os indivíduos em risco em t_i . Nos indivíduos cujo tempo de vida é censurado, estes resíduos são sempre nulos e são indicados como valores omissos. Os resíduos de Schoenfeld são úteis para aferir a permissa de riscos proporcionais subjacente ao modelo de Cox.

Para além da análise gráfica é igualmente possível averiguar a presença de uma correlação linear entre os resíduos e o tempo, através de um teste estatístico. Neste teste são consideradas as seguintes hipóteses:

$$H_0: \rho = 0 \quad vs \quad H_1: \rho \neq 0$$

Assim, admite-se a proporcionalidade dos riscos quando não rejeitamos a hipótese nula de que a correlação é igual a zero.

Resíduos Martingala

O resíduo Martingala, M_i , são baseados em processos de contagem no início do estudo quando todas as covariáveis estão fixas. Este resíduo é definido por:

$$\hat{M}_i = \delta_i - \exp(\hat{\beta}'x_i)\hat{H}_0(t_i) = \delta_i - r_i$$

Assim, M_i é a diferença entre o número de acontecimentos observados no i -ésimo indivíduo e os esperados de acordo com o modelo ajustado (resíduos de Cox-Snell). A análise destes resíduos é útil na identificação de indivíduos mal ajustados ao modelo (por análise gráfica do \hat{M}_i versus índice do indivíduo) e na verificação da adequação da forma funcional das covariáveis incluídas no modelo (interpretação do gráfico do \hat{M}_i do modelo nulo versus covariável, com sobreposição de uma curva de alisamento).

Desvios residuais

Uma das propriedades dos resíduos de martingala é a sua distribuição assimétrica em torno do zero. Assim, foram introduzidos os desvios residuais com o intuito de ultrapassar essa falta de simetria, de modo a facilitar a interpretação dos gráficos dos resíduos. Estes resíduos são dados por:

$$r_{Di} = \text{sgn}(\hat{M}_i) \{-2[\hat{M}_i + \delta_i \log(\delta_i - \hat{M}_i)]\}^{\frac{1}{2}}$$

onde \hat{M}_i é o resíduo martinagla para o i -ésimo indivíduo e $\text{sgn}(\cdot)$ a função sinal. Os desvios residuais são componetes da estatística *Deviance*, dada por $D = -2(\log \hat{L}_c - \log \hat{L}_s)$, onde \hat{L}_c e \hat{L}_s correspondem às verosimilhanças parciais maximizadas do modelo corrente e do modelo saturado, respetivamente. Logo, quanto menor o valor de D , melhor será o ajustamento do modelo. Os gráficos dos desvios residuais *versus* o tempo ou o índice da observação permitem verificar se o modelo é adequado. Para que o modelo seja considerado adequado, os resíduos não devem apresentar qualquer tipo de padrão, devendo ter um comportamento aleatório.

Resíduos score

Os resíduos score permitem analisar a influência de cada observação no ajustamento do modelo e determinar a variância dos coeficientes de regressão. Este resíduo, sob o modelo de Cox, é definido pela diferença entre o valor da covariável x_j correspondente ao i -ésimo indivíduo cujo acontecimento de interesse aconteceu em t_i e a média ponderada de valores dessa covariável para todos os indivíduos, onde o peso correspondente a cada indivíduo em risco é $\exp(\hat{\beta}'x_k)$. Assim, o resíduo score para a j -ésima covariável é dada por

$$S_{ji} = \delta_i \{x_{ji} - \bar{x}_{ji}\}, \text{ para } i = 1, \dots, n; j = 1, \dots, p$$

onde,

$$\bar{x}_{ji} = \frac{\sum_{k \in R(t_i)} x_{jk} \exp(\hat{\beta}'x_k)}{\sum_{k \in R(t_i)} \exp(\hat{\beta}'x_k)},$$

e $R(t_i)$ é o conjunto dos indivíduos em risco no instante t_i . Por último, é de referir que os resíduos associados às observações censuradas tomam o valor zero.

2.5 Modelos de sobrevivência paramétricos

2.5.1 Distribuição exponencial

A distribuição exponencial é caracterizada por ter uma função de risco constante ao longo do tempo. Seja T uma v.a com distribuição exponencial de parâmetro $\lambda > 0$; a função de risco, a função de sobrevivência e a função densidade de probabilidade para $0 \leq t < \infty$, são dadas por:

$$h(t) = \lambda,$$

$$S(t) = \exp(-\lambda t)$$

$$f(t) = \lambda e^{-\lambda t}$$

A função de risco cumulativa é dada por

$$H(t) = -\ln(S(t)) = \lambda t$$

O valor médio e variância são dadas por

$$E(T) = \frac{1}{\lambda}$$

$$var(T) = \frac{1}{\lambda^2}$$

Atendendo a que o tempo de sobrevivência apresenta frequentemente uma distribuição assimétrica, torna-se mais relevante o cálculo da mediana. A mediana é definida como o valor de t para o qual $S(t) = 0,5$. Assim:

$$S(t) = \exp(-\lambda t) = 0,5$$

$$\chi_{1/2} = \frac{\ln(2)}{\alpha}$$

O modelo exponencial é o modelo mais simples, contudo são poucas as situações em que pode ser considerado um risco constante, tendo por isso pouca aplicação prática. No entanto, este modelo pode ser utilizado quando o período de observação é curto, visto que nesse caso é possível assumir que nesse intervalo o risco é constante.

2.5.2 Distribuição de Weibull

A distribuição de Weibull difere da distribuição exponencial porque permite a variação do risco no tempo. Esta distribuição com parâmetro de escala $\lambda > 0$ e parâmetro de forma $\gamma > 0$, tem para $0 \leq t < \infty$,

$$h(t) = \lambda \gamma t^{\gamma-1},$$

$$S(t) = \exp \left\{ - \int_0^t \lambda \gamma u^{\gamma-1} du \right\} = \exp(-\lambda t^\gamma)$$

$$f(t) = \lambda \gamma t^{\gamma-1} \exp(-\lambda t^\gamma)$$

$$H(t) = -\ln S(t) = \lambda t^\gamma$$

No caso particular de $\gamma = 1$ a função de risco assume um valor constante λ , e o tempo de sobrevivência apresenta uma distribuição exponencial. A função de risco

pode ainda ser monótona crescente caso $\gamma > 1$ e monótona decrescente se $0 < \gamma < 1$. O valor médio e a variância são dados por:

$$E(T) = \frac{\Gamma\left(1 + \frac{1}{\gamma}\right)}{\lambda^{1/\gamma}}$$

$$var(T) = \frac{\Gamma\left(1 + \frac{2}{\gamma}\right) - \Gamma^2\left(1 + \frac{1}{\gamma}\right)}{\lambda^{1/\gamma}}$$

A mediana da distribuição é dada por

$$S(t) = \exp(-\lambda t^\gamma) = 0,5$$

$$\chi_{1/2} = \left(\frac{\ln(2)}{\lambda}\right)^{1/\gamma}$$

A distribuição de Weibull é atualmente uma das mais utilizadas na análise de tempos de sobrevivência na área da saúde, visto ser razoavelmente flexível, pois permite a variação do risco no tempo.

2.5.3 Distribuição log-logística

Na distribuição log-logística, com um parâmetro de escala $\lambda > 0$ e parâmetro forma $\alpha > 0$, para $0 \leq t < \infty$, tem-se que

$$h(t) = \frac{\alpha \lambda t^{\alpha-1}}{1 + \lambda t^\alpha}$$

$$S(t) = \frac{1}{1 + \lambda t^\alpha}$$

$$f(t) = \frac{\alpha \lambda t^{\alpha-1}}{(1 + \lambda t^\alpha)^2}$$

que representam a função de risco, a função de sobrevivência e a função densidade de probabilidade. O modelo log-logístico pode ser uma alternativa ao modelo de Weibull nas situações em que este não é adequado, visto permitir considerar um risco não monótono unimodal. Neste caso, a função de risco é monótona decrescente se $0 < \alpha \leq 1$, sendo unimodal caso $\alpha > 1$.

A mediana da distribuição é dada por:

$$\chi_{1/2} = \lambda^{-1/\alpha}$$

2.5.4 Distribuição log-normal

Se assumirmos que o tempo de sobrevivência T segue uma distribuição log-normal, isto é, que o logaritmo de T segue uma distribuição normal com valor médio μ e variância σ^2 , então a função de risco, a função de sobrevivência e a função densidade de probabilidade são dadas por

$$h(t) = \frac{\frac{1}{\sqrt{2\pi}\sigma t} \exp\left[-\frac{1}{2}\left(\frac{\ln(t) - \mu}{\sigma}\right)^2\right]}{1 - \Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)}$$

$$S(t) = 1 - \Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)$$

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma t} \exp\left[-\frac{1}{2}\left(\frac{\ln(t) - \mu}{\sigma}\right)^2\right]$$

para $t > 0$, onde $\Phi(\cdot)$ é a função de distribuição da $N(0,1)$. Tal como o modelo log-logístico, este modelo tem uma função de risco unimodal. A função de risco deste modelo é tal que $h(0) = 0$, sendo crescente até um valor máximo e depois decrescente. O instante em que ocorre o valor máximo depende do valor de σ . O valor médio e a variância da v.a. T são expressas da seguinte forma:

$$E(T) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

$$var(T) = [\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2)$$

A mediana é dada por:

$$\chi_{1/2} = \exp(\mu)$$

2.5.5 Distribuição Gama

A distribuição gama com parâmetro de escala $\lambda > 0$ e parâmetro forma de $\alpha > 0$, tem uma f.d.p. dada por

$$f(t) = \frac{\lambda(\lambda t)^{\alpha-1} e^{-(\lambda t)}}{\Gamma(\alpha)}$$

para $t \geq 0$. Não existe uma expressão analítica fechada para a função de sobrevivência, sendo esta dada por

$$S(t) = 1 - I(\alpha, \lambda t),$$

onde $I(\alpha, x)$ é a função gama incompleta. A função de risco pode ser monótona crescente (se $\alpha > 1$), monótona decrescente (se $0 < \alpha < 1$) e constante (se $\alpha = 1$, sendo neste caso a distribuição exponencial). De notar que também não existe uma fórmula explícita para a função de risco; no entanto, pode ser facilmente obtida a partir da razão entre a f.d.p e a função de sobrevivência, $\lambda(t) = f(t)/S(t)$. O valor médio de T e a variância são expressos da seguinte forma:

$$E(T) = \frac{k}{\lambda}$$

$$var(T) = \frac{k}{\lambda^2}$$

2.5.6 Distribuição de Gompertz

A distribuição de Gompertz é importante na descrição do padrão de mortalidade nos adultos, visto que esta distribuição apresenta uma função de risco exponencial crescente, que é compatível com a mortalidade humana. Nesta distribuição, tem-se que

$$h(t) = \theta \exp(\alpha t)$$

$$S(t) = \exp\left\{\frac{\theta}{\alpha} [1 - \exp(\alpha t)]\right\}$$

$$f(t) = \theta \exp(\alpha t) \exp\left\{\frac{\theta}{\alpha} [1 - \exp(\alpha t)]\right\}$$

para $t \geq 0$ e $\theta > 0$. A função de risco é monótona crescente quando $\alpha > 0$; e monótona decrescente para $\alpha < 0$.

A mediana é dada por:

$$\chi_{1/2} = \frac{1}{\alpha} \ln\left(\frac{\alpha}{\theta} \ln 2 + 1\right)$$

2.5.7 Função de verosimilhança

Na construção da função de verosimilhança é importante considerar que o tempo de censura é independente do tempo até à ocorrência do acontecimento de interesse. Seja n o número de indivíduos em estudo, e que os dados relativos ao i -ésimo indivíduo estão na forma (t_i, δ_i, x_i) , sendo $i = 1, \dots, n$; t_i o tempo de vida observado ($\delta_i = 1$) ou censurado ($\delta_i = 0$) e x_i um vetor de covariáveis fixas. Suponhamos que a distribuição do tempo de vida T dado x é conhecida com a exceção do vetor de parâmetros θ , sobre o qual se pretende fazer inferência, e que a função de sobrevivência é dada por $S(t_i; x_i, \theta)$ e a função densidade de probabilidade é $f(t_i; x_i, \theta)$. Assim, a função de verosimilhança é dada por:

$$L(\theta) = \prod_{i=1}^n f(t_i; x_i, \theta)^{\delta_i} S(t_i; x_i, \theta)^{1-\delta_i}$$

Sabendo a distribuição do tempo de vida o passo seguinte será substituir, na função de verosimilhança a respetiva função densidade de probabilidade e função de sobrevivência. Nalgumas distribuições, como é o caso da distribuição de Weibull ou log-normal, as equações não têm uma solução explícita para todos os parâmetros, pelo que são utilizados métodos iterativos para estimá-los, que é habitualmente feito pelo *software* estatístico.

2.6 Modelos de regressão paramétricos

Atendendo a que a sobrevivência pode ser influenciada por características inerentes aos indivíduos, os modelos de regressão paramétricos permitem incorporar covariáveis para que estas sejam consideradas no tempo de sobrevivência. Assim, o tempo de sobrevivência será a variável dependente, ao passo que as covariáveis serão consideradas variáveis independentes. Os modelos anteriormente apresentados podem ser adaptados de modo a incluir a influência das covariáveis através de um vetor de covariáveis, $x = (x_1, \dots, x_p)'$, e de parâmetros $\beta = (\beta_1, \beta_2, \dots, \beta_k)'$. O vetor de covariáveis dado por $x = 0$ é definido para um conjunto de circunstâncias padrão.

O modelo de Cox é o modelo mais utilizado dada a maior facilidade na aplicação e interpretação dos resultados, sendo também mais flexível do que os modelos paramétricos, visto não ser necessário definir a função de risco padrão. No modelo de Cox não é necessário especificar a distribuição dos tempos de sobrevivência, contudo é essencial verificar o pressuposto de riscos proporcionais para que os resultados sejam fiáveis. No entanto, é de salientar que no caso do modelo de Cox ser adequado, caso seja possível definir a distribuição dos tempos de sobrevivência, os modelos

paramétricos são mais informativos. Em contrapartida, a principal limitação dos modelos paramétricos é a necessidade de determinar a distribuição mais apropriada para os tempos de sobrevivência.

Os modelos de regressão paramétricos mais comuns são divididos em modelos de riscos proporcionais, tempo de vida acelerado e de possibilidades proporcionais. Cada um destes grupos representa uma família de modelos, que partilham características semelhantes na caracterização da relação entre o tempo de sobrevivência e as covariáveis.

2.6.1 Modelos de tempo de vida acelerado

A função de risco para um indivíduo com covariáveis x é dada por:

$$h(t; z) = h_0(t) \exp(\beta'x)$$

No caso dos modelos de tempo de vida acelerado as covariáveis irão acelerar ou tornar mais lento o tempo de vida, visto terem um efeito multiplicativo no tempo de sobrevivência. Também neste caso podemos ter uma abordagem paramétrica ou semi-paramétrica, consoante a função $h_0(t)$ seja especificada ou não.

Os modelos paramétricos de tempo de vida acelerado têm em comum a representação log-linear. O modelo log linear é dado por

$$\log T = \mu + \alpha'x + \sigma\varepsilon$$

em que μ é o termo independente, α o vetor de parâmetros de regressão e σ um parâmetro de escala. A quantidade ε é uma variável aleatória incluída para modelar o desvio dos valores de $\log T$ da parte linear do modelo, sendo assumida uma determinada distribuição para ε . A justificação para a designação deste modelo, como modelo de tempo de vida acelerado pode ser demonstrada. Consideremos a v.a. $T_0 = \exp(\mu + \sigma\varepsilon)$ cuja função de sobrevivência é $S_0(t) = P[\exp(\mu + \sigma\varepsilon) > t]$. Assim:

$$\begin{aligned} S(t; x) &= P(T > t | x) \\ &= P[\exp(\mu + \alpha'x + \sigma\varepsilon) > t] \\ &= P[\exp(\mu + \sigma\varepsilon) > t \exp(-\alpha'x)] \\ &= S_0(t \exp(-\alpha'x)) = S_0(t / \exp(\alpha'x)) \end{aligned}$$

De acordo com esta expressão, observa-se que o efeito das covariáveis consiste na modificação da escala tempo através do fator $\exp(-\alpha'x)$, que é normalmente

designado por fator de aceleração. Efetivamente, o tempo de sobrevivência de um indivíduo a que está associado um vetor de covariáveis x é:

$$T = T_0 / \exp(-\alpha'x)$$

- Se $\exp(-\alpha'x) > 1 \Leftrightarrow \alpha'x < 0$, o tempo até à ocorrência do acontecimento de interesse é acelerado pelo efeito das covariáveis.
- Se $\exp(-\alpha'x) < 1 \Leftrightarrow \alpha'x > 0$, o tempo até à ocorrência do acontecimento de interesse é travado pelo efeito das covariáveis.

Por último, é de notar que como consequência de $S(t; x) = S_0(t \exp(-\alpha'x))$, a mediana do tempo de sobrevivência de um indivíduo com vetor de covariáveis x é igual à mediana do tempo de sobrevivência do indivíduo padrão multiplicada pelo inverso do fator de aceleração.

2.6.2 Modelo de riscos proporcionais

A função de risco de T , dado x , é dada por:

$$h(t; x) = h_0(t)\varphi(x)$$

em que $h_0(t)$ representa a função de risco do indivíduo com vetor de covariáveis $x = 0$. Como referido anteriormente a função $\varphi(x)$ pode ser parametrizada por $\varphi(x; \beta) = \exp(\beta'x)$. O fator de proporcionalidade $\varphi(x)$ corresponde ao risco relativo, obtido pelo quociente entre o risco de morte de um indivíduo associado a um vetor de covariáveis x e o risco de morte para um indivíduo com vetor de covariáveis $x = 0$. É de notar que neste caso as covariáveis têm um efeito multiplicador na função de risco do indivíduo padrão. A função de sobrevivência de T , dado x , é dada por:

$$S(t; z) = [S_0(t)]^{\varphi(x)}$$

onde S_0 é a função de sobrevivência associada à função de risco do indivíduo padrão. A razão de ser denominado um modelo de regressão de riscos proporcionais, tem a ver com o facto de que o quociente das funções de risco de dois indivíduos com vetores de covariáveis x_1 e x_2 , dado por $h(t; x_1)/h(t; x_2)$ não depende de t . A diferença entre este modelo e o modelo de tempo de vida acelerado é que no último caso, o risco pode ser variável ao longo do tempo. Por último, é de salientar que quando $h_0(t)$, é especificada temos um modelo paramétrico, caso contrário estaremos perante uma abordagem semi-paramétrica proposta por Cox.

2.6.3 Modelos de possibilidades proporcionais

Neste modelo a possibilidade de um indivíduo com vetor de covariáveis x , sobreviver para além do instante t é dada por

$$\frac{S(t)}{1 - S(t)} = e^{\eta} \frac{S_0(t)}{1 - S_0(t)}$$

onde $\eta = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ e x_j representa o valor da j -ésima covariável, $j = 1, \dots, p$ e $S_0(t)$ a função de sobrevivência para o indivíduo padrão. Neste modelo as variáveis têm um efeito multiplicativo na possibilidade (*odds*) de um indivíduo sobreviver para além do instante t . O logaritmo da razão entre a possibilidade de um indivíduo sobreviver para além de um instante t e um indivíduo padrão, é η . Assim sendo, trata-se de um modelo linear para o logaritmo da razão de possibilidades.

De forma análoga ao modelo de riscos proporcionais pode ser obtida uma estimativa não paramétrica da função de risco. O ajustamento do modelo é conseguido com base nos dados, através da estimativa dos parâmetros β da componente linear do modelo e da função de sobrevivência. No entanto, quando é especificada uma distribuição para os tempos de sobrevivência estamos perante uma versão paramétrica dos modelos de possibilidades proporcionais. A função de risco é dada por:

$$\frac{h(t; z)}{h_0(t)} = \{1 + (e^{\eta} - 1)S_0(t)\}^{-1}.$$

É de notar que à medida que t aumenta de 0 a ∞ , a função de sobrevivência é monótona decrescente. Quando $S_0(t) = 1$, $\frac{h(t; z)}{h_0(t)} = e^{-\eta}$, e quando $t \rightarrow \infty$, a razão das funções de risco converge para a unidade. Na prática, este modelo é adequado quando é esperada uma convergência das funções de risco de dois ou mais grupos de doentes sujeitos a tratamentos diferentes, por exemplo quando a cura da doença em estudo seja uma possibilidade.

2.6.4 Modelo de regressão de Weibull

O modelo Weibull é concomitantemente um modelo de riscos proporcionais e um modelo de tempo de vida acelerado. No primeiro caso a função de risco de um indivíduo com vetor de covariáveis x é dada por

$$h(t; x) = h_0(t) \exp(\beta' x) = \lambda \gamma t^{\gamma-1} \exp(\beta' x),$$

com parâmetro de escala $\lambda \exp(\beta'x)$ e parâmetro de forma γ . Este resultado é uma manifestação da propriedade de riscos proporcionais, visto que as covariáveis alteram o parâmetro de escala, enquanto o parâmetro de forma irá manter-se constante. A correspondente função de sobrevivência é dada por:

$$S(t; x) = \exp(-\lambda t^\gamma \exp(\beta'x)).$$

No entanto, sendo o modelo de Weibull também um modelo de tempo de vida acelerado, permite uma representação log-linear. Neste caso, o logaritmo do tempo de vida pode ser escrito da seguinte forma:

$$\log T = \mu + \alpha'x + \sigma\varepsilon.$$

Se T apresenta uma distribuição de Weibull, ε segue uma distribuição de Gumbel com função densidade de probabilidade $f(x) = \exp(x - \exp(x))$, para $-\infty < x < +\infty$. A função de sobrevivência é dada por:

$$S(t; x) = \exp\left[-\exp\left(\frac{\log t - \mu - \alpha'x}{\sigma}\right)\right] = \exp\left[-\exp(-\mu/\sigma)t^{1/\sigma} \exp((-\alpha/\sigma)'x)\right].$$

A mediana pode ser obtida a partir da seguinte expressão

$$\mu \log(2)^{1/\alpha}$$

Por comparação da função de sobrevivência obtida a partir do modelo de riscos proporcionais e a função de sobrevivência resultante da representação log-linear correspondente ao modelo de tempo de vida acelerado, verificamos as seguintes relações entre os parâmetros do modelo Weibull:

$$\lambda = \exp(-\mu/\sigma), \gamma = 1/\sigma \text{ e } \beta_j = -\alpha/\sigma.$$

Estas relações são muito importantes visto que habitualmente os *softwares* de estatística, apenas fornecem os parâmetros relativos ao modelo de tempo de vida acelerado.

2.6.5 Modelo de regressão log-logístico

Este modelo é o único que é concomitantemente um modelo de possibilidades proporcionais e um modelo de tempo de vida acelerado. A função de sobrevivência, sob o modelo de possibilidades proporcionais para um indivíduo com vetor de variáveis x é dado por:

$$S(t; x) = \frac{1}{1 + \lambda \exp(\beta'x)t^k}$$

Assim, o tempo de sobrevivência desse indivíduo segue uma distribuição log-logística com parâmetro de escala $\lambda \exp(\beta'x)$ e parâmetro de forma k . É de notar que:

$$\frac{S(t; x)}{1 - S(t; x)} = \exp(\beta'x) \frac{S_0(t)}{1 - S_0(t)}$$

No modelo log-logístico sob o modelo de tempo de vida acelerado considerando a representação log-linear, $\log T_i = \mu + \alpha'x + \sigma\varepsilon$, ε segue uma distribuição logística. A função de risco e a função de sobrevivência para o i -ésimo indivíduo é dada por

$$h(t) = \frac{e^{\theta - k\eta} k t^{k-1}}{1 + e^{\theta - k\eta} t^k}$$

$$S(t) = \left[1 + \exp \frac{\log t - \mu - \alpha'z}{\sigma} \right]^{-1}$$

onde $\lambda = \exp(-\mu/\sigma)$, $k = 1/\sigma$ e $\beta_j = -\alpha_j/\sigma$.

2.6.6 Critério de Informação de Akaike (AIC)

A principal limitação dos modelos paramétricos tem a ver com a necessidade de definir a forma da função de risco. Considerando modelos não aninhados o AIC pode ser utilizado para escolher o modelo mais adequado. Este critério irá penalizar a log verosimilhança de cada modelo, de forma a refletir o número de parâmetros que estão a ser estimados, permitindo a comparação de modelos. A expressão relativa ao AIC é dada por

$$AIC = -2 \ln \hat{L} + 2(k + c)$$

onde k é o número de covariáveis e c é o número de parâmetros específicos de cada modelo. De um modo geral, é feita a comparação dos valores de AIC calculados para os diferentes modelos paramétricos e escolhe-se o modelo com o valor de AIC mais baixo.

Distribuição	Forma da função de risco	c
Exponencial	Constante	1
Weibull	Monótona	2
Gompertz	Monótona	2
Log-normal	Variável	2
Log-logística	Variável	2

Tabela 1: Comparação de modelos paramétricos.

2.6.7 Análise de Resíduos

A interpretação da análise de resíduos para modelos paramétricos é idêntica à que foi descrita anteriormente para o modelo de Cox. Contudo, nesta seção será abordada a análise de resíduos no contexto dos modelos paramétricos.

Resíduos de Cox-Snell

Os resíduos de Cox-Snell são usualmente utilizados para averiguar o ajustamento global do modelo de Cox, no entanto são igualmente aplicáveis nos modelos de tempo de vida acelerado. Estes resíduos são definidos por

$$r_i = \hat{H}(t_i; x_i)$$

em que para o i -ésimo indivíduo, se tem que t_i é o instante, x_i o vetor de covariáveis e \hat{H} a função de risco cumulativo estimado para o modelo. Considerando o ajustamento de um modelo paramétrico, cuja representação linear é dada por :

$$\ln T_i = \mu + \gamma_i' x_i + \sigma \varepsilon_i$$

sendo $i = 1, \dots, n$. Se assumirmos que as covariáveis x_i são fixas, $S(T_i | x_i)$ apresenta uma distribuição Uniforme em $(0,1)$, logo a v.a.

$$r_i = H(T_i | x_i) = -\ln [S(T_i | x_i)] = \int_0^{T_i} h(u | x_i) du$$

apresenta uma distribuição exponencial de valor médio 1. Considerando um modelo de tempo de vida acelerado, a função de risco pode ser dada por:

$$h(t|z_i) = \exp(\beta' z_i) h_0[t \exp \beta' z_i] .$$

Logo, os resíduos de Cox-Snell para um modelo de tempo de vida acelerado são dados por:

$$\hat{r}_i = \int_0^{t_i} \exp(\hat{\beta}' z_i) \hat{h}_0[u \exp(\hat{\beta}' z_i)] du = \hat{H}_0[t_i \exp(\hat{\beta}' z_i)] = -\ln [\hat{S}_0(t_i \exp(\hat{\beta}' z_i))]$$

Assim, sabendo que r_i segue uma distribuição exponencial (1), se o modelo for adequado a v.a. \hat{r}_i irá comportar-se como uma amostra censurada proveniente de uma população exponencial unitária. E tal como foi apresentado para o modelo de Cox, é então possível verificar se o modelo é adequado com base na representação gráfica dos resíduos de Cox-Snell *versus* a estimativa de Nelson-Aalen para a função de risco cumulativa para os resíduos, sendo esperada uma linha reta com declive aproximadamente 1 e ordenada na origem nula.

Por último é importante ter em conta que esta análise não nos permite identificar qual o problema no ajustamento quando o gráfico não é linear. Para além disso, é de ter em conta que a distribuição exponencial dos resíduos é observada, quando utilizados os verdadeiros valores dos parâmetros do modelo e não quando utilizadas as estimativas desses parâmetros.

Resíduos Martingala

Os resíduos Martingala para modelos paramétricos onde todas as covariáveis são fixas, são definidos por:

$$\hat{M}_i = \delta_i - \hat{r}_i$$

Sendo que \hat{r}_i , representa os resíduos de Cox-Snell e δ_i é igual a 1 quando t_i é um tempo de vida observado, e é zero quando t_i corresponde a um tempo de vida censurado. Os resíduos martingala são não correlacionados e têm valor esperado zero; não se dispõem de forma simétrica em torno do valor médio, são negativos para as observações censuradas, e tomam valores entre $-\infty$ e 1. De um modo simplista, os resíduos martingala permitem analisar o excesso de mortes não previstas pelo modelo. É importante referir que para modelos paramétricos, a derivação de \hat{M}_i como martingala não é válida, contudo o nome mantém-se atendendo a que a interpretação dos resíduos é idêntica ao que foi descrito para o modelo de Cox.

Desvios residuais

Os desvios residuais que foram descritos para o modelo de Cox, têm igual aplicação nos modelos de regressão paramétricos. Estes resíduos foram propostos de modo a ultrapassar a assimetria característica dos resíduos Martingala e assim facilitar a interpretação dos gráficos. Os desvios residuais permitem verificar de uma forma global a qualidade do ajustamento do modelo, visto que pela análise dos *outliers* é possível observar se o acontecimento de interesse ocorre “muito cedo” ou “muito tarde” em relação ao que é predito pelo modelo ajustado.

Capítulo 3-Análise do tempo até à ocorrência de excesso de peso/obesidade

3.1 Descrição do estudo

Trata-se de um estudo observacional que incluiu doentes submetidos a TCPH-Alo, na Unidade de Transplante de Medula do Instituto Português de Oncologia de Lisboa-Francisco Gentil E.P.E. (IPOL-FG). Os dados foram colhidos retrospectivamente a partir de 31 de Maio de 2011, pelo que esta foi considerada a data limite do estudo. Foram registados dados de doentes transplantados entre 28 de Maio de 1987 e 31 de Dezembro de 2007, para que todos os doentes tivessem um tempo de seguimento clinicamente relevante. Embora tenha sido recolhida informação de 103 doentes sujeitos a transplante alogénico, a análise estatística dos dados foi limitada a 83 doentes, uma vez que 20 foram excluídos por não ter sido possível obter toda a informação necessária ou pela imunossupressão ter sido mantida até à última observação.

A definição do acontecimento de interesse foi feita com base no Índice de Massa Corporal (IMC), que permite classificar o estado nutricional do doente de acordo com a seguinte escala: magreza ($IMC < 18.5$), peso normal ($18.5-24.9$), excesso de peso ($25-29.9$), obesidade tipo I ($30-34.9$), obesidade tipo II ($35-39.9$) e obesidade mórbida (≥ 40). O IMC é calculado com base na expressão:

$$IMC = \frac{Peso(kg)}{Altura(m)^2}$$

Especificamente definiu-se o acontecimento de interesse como o aumento do IMC que provoque a transição do indivíduo para uma categoria superior correspondente a excesso de peso ou obesidade. Após o TCPH-Alo, os doentes são seguidos na consulta médica com periodicidade concordante com a sua necessidade individual. Depois de estabilizada a situação clínica, os doentes são reavaliados anualmente.

A análise de sobrevivência foi utilizada para investigar quais os fatores que têm influência significativa no tempo decorrido até ao acontecimento de interesse, que é a variável resposta. Para cada indivíduo foi considerado como instante inicial a data de paragem da imunossupressão (com ciclosporina, tacrolimus, moftil ou cotricóides) e foi

registado o tempo decorrido até à data da ocorrência do acontecimento de interesse ou a última data no processo. Assim sendo, para os indivíduos para os quais foi registado a alteração do IMC para uma categoria superior correspondente a excesso de peso ou obesidade, registou-se o valor observado da variável resposta e os restantes indivíduos deram origem a observações censuradas. A decisão de iniciar a contagem do tempo após a paragem da imunossupressão é justificada pela influência deste tratamento no aumento de peso à custa da retenção de líquidos.

3.2 Definição das variáveis

Com base na informação recolhida na folha de registo de dados (Anexo 1), foram definidas as variáveis constantes nas seguintes tabelas 2 e 3.

Variáveis	Código	Tipo
Género	0-Feminino	Dicotómica
	1-Masculino	
Diagnóstico	1-Leucemia Linfóide Aguda (LLA)	Categórica
	2-Leucemia Mieloide Crónica (LMC)	
	3-Leucemia Mieloide Aguda (LMA)	
	4-Outros diagnósticos	
<i>Total Body Irradiation</i>	0-Não	Dicotómica
	1-Sim	
Dose (Gy)		Contínua

Tabela 2:Definição das variáveis.

Variáveis	Código	Tipo
Ciclosporina	0-Não	Dicotómica
	1-Sim	
Tacrolimus	0-Não	Dicotómica
	1-Sim	
Corticóide	0-Não	Dicotómica
	1-Sim	
Mofetil	0-Não	Dicotómica
	1-Sim	
Idade no transplante (anos)	---	Contínua
Tempo de internamento (dias)	---	Contínua
Diabetes	0-Não	Dicotómica
	1-Sim	
Hipertensão	0-Não	Dicotómica
	1-Sim	
Dislipidemia	0-Não	Dicotómica
	1-Sim	
<i>Graft vs Host Disease</i>	0-Não	Dicotómica
	1-Sim	
Peso admissão (kg)	---	Contínua
Peso na alta (kg)	---	Contínua
Peso na paragem da imunossupressão	---	Contínua
Peso na última observação (kg)	---	Contínua
Altura (m)	---	Contínua

Tabela 3:Definição das variáveis (continuação).

3.3 Análise preliminar

Para as variáveis contínuas a comparação entre os géneros foi realizada com testes paramétricos (teste t para amostras independentes) ou não paramétricos (teste de Mann-Whitney) conforme o ajustamento da distribuição normal aos dados, se mostrou adequado ou não pelo teste de Kolmogorov-Smirnov. A existência de associação entre o género e cada uma das variáveis qualitativas foi analisada com o teste do qui-quadrado de independência. Na análise da diferença entre o peso médio entre a paragem da imunossupressão e o peso médio na última observação, foi utilizado o teste t para amostras emparelhadas. Estas análises foram realizadas com o *software* estatístico SPSS versão 20. Foram igualmente construídos gráficos *boxplot* e histogramas para caracterização da variável peso nos quatro momentos em que foi avaliada.

Foram incluídos no estudo um total de 83 doentes, 36 mulheres (43,4%) e 47 homens (56,6%), sendo que as mulheres apresentaram uma idade média na data do transplante significativamente superior à dos homens. No entanto, não foi observada uma associação estatisticamente significativa entre o género e cada uma das restantes variáveis. Os dados clínicos encontram-se resumidos na tabela 4 e 5, onde está também indicado o valor-p correspondente aos testes estatísticos anteriormente mencionados.

	Feminino (n=36)	Masculino (n=47)	valor-p
Idade transplante (anos)	37.6±9.45	33.4±9.91	0.042
Tempo de internamento (dias)	36.5±16.8	39.3±24.4	0.738
Diagnóstico:			
Leucemia Linfóide Aguda	5(45.5)	6(54.5)	0.688
Leucemia Mielóide Crónica	17(50.0)	17(50.0)	
Leucemia Mielóide Aguda	7(43.8)	9(56.2)	
Outros diagnósticos	7(33.3)	14(66.7)	
TBI			
Sim	4(26.7)	11(73.3)	0.149
Não	32(47.1)	36(52.9)	

Tabela 4:Caracterização da amostra estudada: resultados expressos em média ±desvio-padrão ou número de doentes (percentagem). Valor-p associado aos testes.

	Feminino (n=36)	Masculino (n=47)	valor-p
Ciclosporina			
Sim	33(43.4)	43(56.6)	
Não	3(42.9)	4(57.1)	0.977
Tacrolimus			
Sim	32(44.4)	40(55.6)	
Não	4(36.4)	7(63.6)	0.614
Corticóide			
Sim	16(50.0)	16(50.0)	
Não	20(39.2)	31(60.8)	0.335
Mofetil			
Sim	32(46.4)	37(53.6)	
Não	4(28.6)	10(71.4)	0.220
Diabetes			
Sim	2(66.7)	1(33.3)	
Não	34(42.5)	46(57.5)	0.407
Hipertensão			
Sim	10(50.0)	10(50.0)	
Não	26(41.3)	37(58.7)	0.492
Dislipidemia			
Sim	2(28.6)	5(71.4)	
Não	34(44.7)	42(55.3)	0.409
GVHD			
Sim	16(34.4)	28(63.6)	
Não	20(52.6)	18(47.4)	0.139

Tabela 5: Caracterização da amostra estudada (continuação): resultados expressos em média \pm desvio-padrão ou número de doentes (percentagem). Valor-p associado aos testes.

Pela observação das *boxplot* apresentadas na figura 4, relativas ao peso em todos os momentos de pesagem, os homens apresentam uma mediana do peso superior à das mulheres, sendo no entanto de notar que na última medição esta diferença não é tão marcada. A variabilidade dos pesos entre as mulheres e entre os homens parece ser idêntica nos diferentes momentos de avaliação, com a exceção do momento da paragem da imunossupressão (as mulheres apresentam maior variabilidade do que os homens). Os histogramas apresentados na figura 5 mostram que em todos os momentos de pesagem parece existir assimetria positiva na distribuição dos pesos.

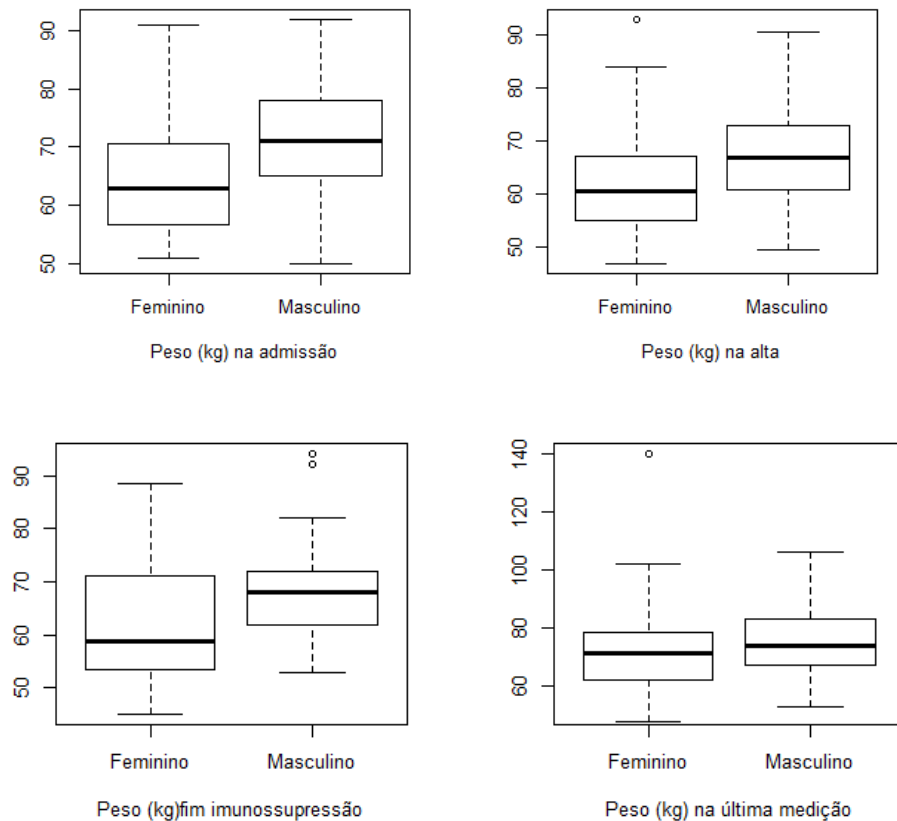


Figura 4: Distribuição do peso por género em quatro momentos de avaliação (admissão, alta, paragem da imunossupressão e última observação).

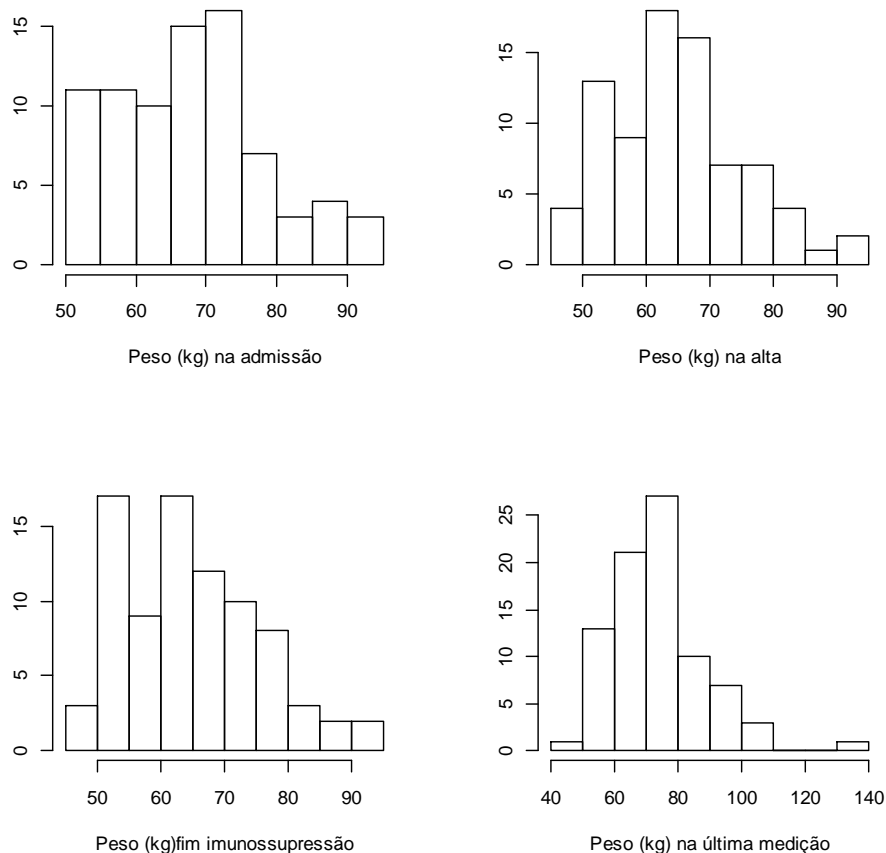


Figura 5: Histograma da variável peso nos diferentes momentos de avaliação (admissão, alta, paragem da imunossupressão).

Neste estudo constatou-se ainda que em média os homens são significativamente mais altos e mais pesados do que as mulheres. A diferença relativa ao peso manteve-se significativa em todas as medições de peso registadas (peso alta e paragem da imunossupressão), com a exceção do último peso registado, na qual deixa de existir uma diferença estatisticamente significativa entre homens e mulheres. É de salientar que foi observado um aumento significativo na média do peso entre a paragem da imunossupressão e o último registo. No que diz respeito ao IMC verificou-se que as mulheres apresentaram um valor superior ao dos homens, sendo a média de IMC referente ao último registo significativamente superior ao IMC na paragem da imunossupressão. A tabela 6 resume os resultados obtidos na análise do peso e IMC. É de salientar que globalmente a prevalência de excesso de peso e obesidade na última observação foi superior à inicial (excesso de peso: 18.3% vs. 43.9%; obesidade 7.3% vs. 18.3%).

	Feminino n=36	Masculino n=47	valor-p	valor-p
Altura (m)	1.58±0.06	1.72±0.07	<0.001	
Peso admissão (kg)	64.2±9.98	71.03±9.96	0.003	
Peso alta (kg)	62,2±10.53	67.2±9.4	0.010	
Peso paragem imunossupressão (kg)	62.1±11.30	67.9±9.11	0.005	<0.001*
Peso última observação (kg)	71.5±17.6	71.3±11.9	0.183	
IMC na paragem da imunossupressão	24.7±4.12	22.7±2.95	0.027	
IMC na admissão				
Magreza	1(33.3)	2(66.7)	0.282	
Peso normal	22(37.9)	36(62.1)		
Excesso de peso	9(60.0)	6(40.0)		
Obesidade	4(66.7)	2(33.3)		
IMC na última observação	28.5±7.3	24.9±3.7	0.013	<0.001**
Magreza	0(0)	1(100)	0.003	
Peso normal	9(30.0)	21(70.0)		
Excesso de peso	14(38.9)	22(61.1)		
Obesidade	13(86.7)	2(13.3)		

Tabela 6: Caracterização da evolução ponderal. Resultados expressos em média ± desvio-padrão ou número de doentes (percentagem). *Comparação entre o peso na paragem da imunossupressão e o último peso registado; **Comparação entre o IMC na paragem da imunossupressão e o último na última observação.

3.4 Estimação não paramétrica

É de salientar que a análise de sobrevivência foi efetuada para um total de 78 indivíduos sendo 34 mulheres (43.6%) e 44 homens (56.4%), visto que foram excluídos 5 dos 83 indivíduos por apresentarem dados incompletos. O acontecimento de interesse foi observado em 31 (39.7%) indivíduos, sendo o número de observações censuradas igual a 47 (60.3%) e a mediana do tempo de seguimento de 7.6 anos (0.03-18.95).

A estimativa da função de sobrevivência foi obtida com base no estimador de Kaplan-Meier e optou-se pelo intervalo de confiança com transformação Log-Log (figura 6 e Anexo II). A estimativa da mediana do tempo até à ocorrência do acontecimento de interesse, obtida com base no estimador de Kaplan-Meier, foi de 15.3 anos. A estimativa da mediana do tempo até ao acontecimento de interesse foi de 15.3 anos nas mulheres. Nos homens não foi possível estimar a mediana, visto que a estimativa da função de sobrevivência é superior a 0.5 para todos os instantes.

É importante referir que neste caso a estimativa de Kaplan-Meier em cada instante t representa a percentagem estimada de indivíduos que ainda não subiram no IMC. A figura 7 apresenta a estimativa de Kaplan-Meier da função de sobrevivência para cada variável estudada. Nesta figura é possível observar que na maioria das variáveis existe uma sobreposição das curvas de sobrevivência, com a exceção das variáveis Idade no transplante, IMC no transplante, ciclosporina, dislipidemia e diabetes. É de notar que para a idade no transplante as curvas de sobrevivência são praticamente coincidentes até aproximadamente 4 anos após o transplante. A partir daí, estima-se que a percentagem de indivíduos que ainda não subiram no IMC é sempre superior nos indivíduos transplantados com idade maior ou igual a 35 anos. Da mesma maneira, parece existir um melhor prognóstico para doentes que apresentam no transplante um IMC superior ou igual a 24.9.

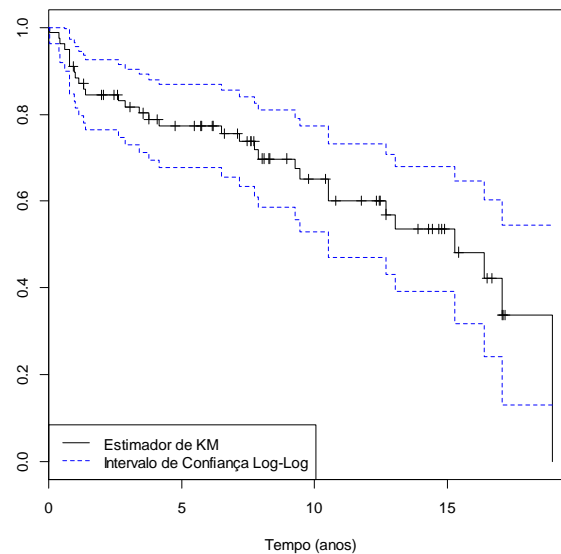
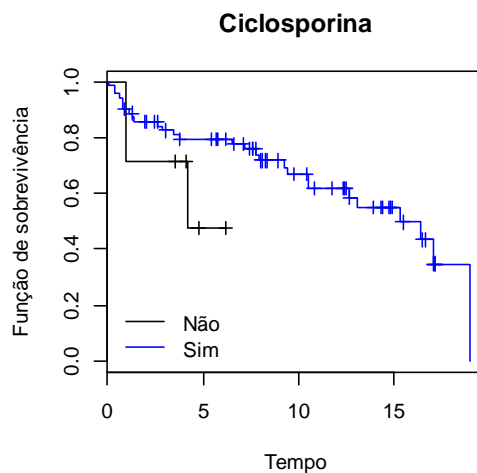
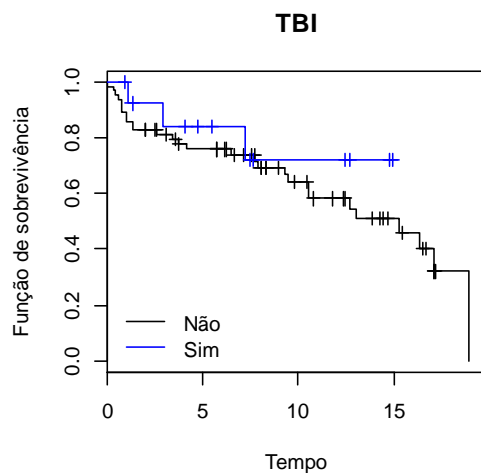
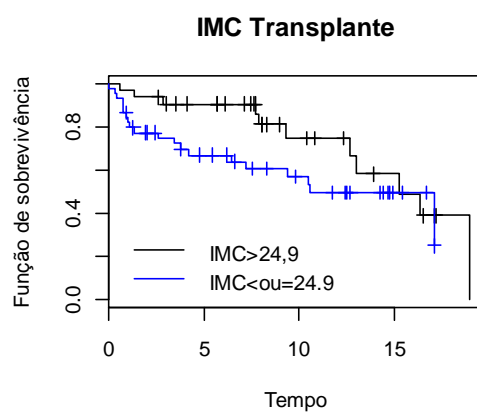
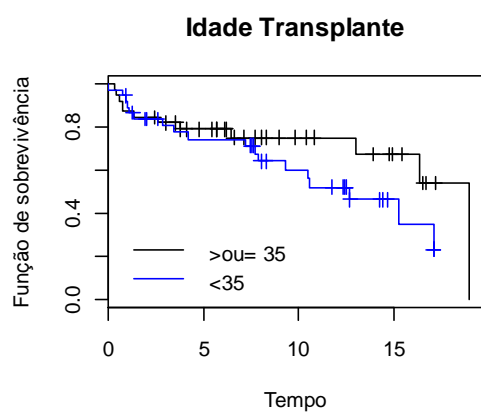
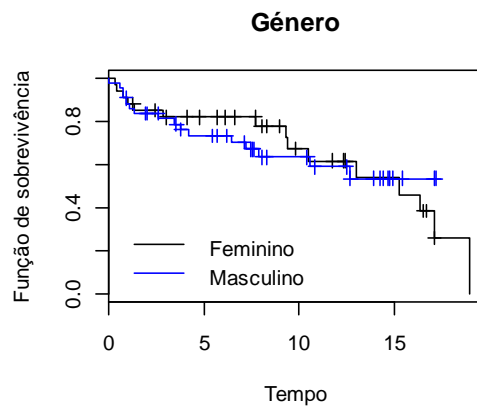
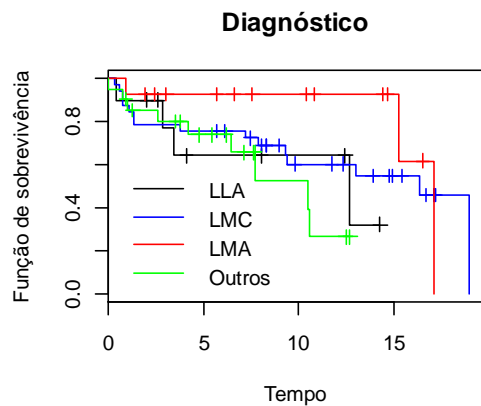
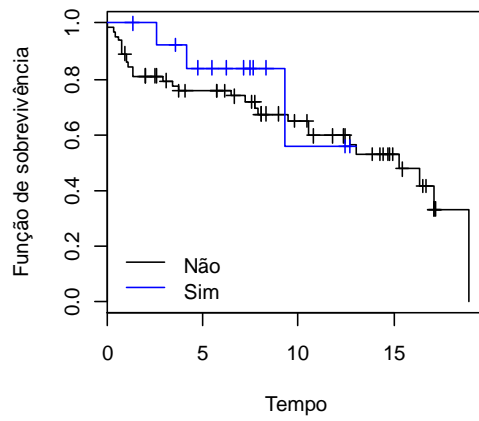


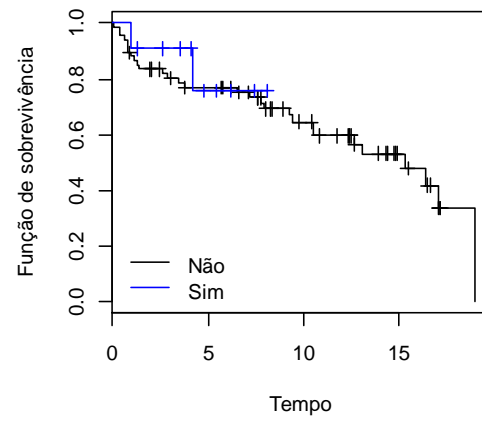
Figura 6: Estimativa de Kaplan-Meier da função de sobrevivência e intervalos de confiança com transformação Log-Log.



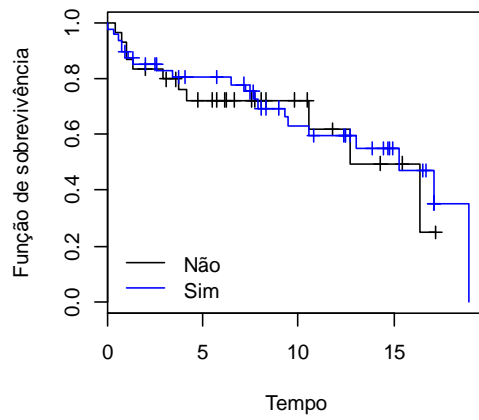
Mofetil



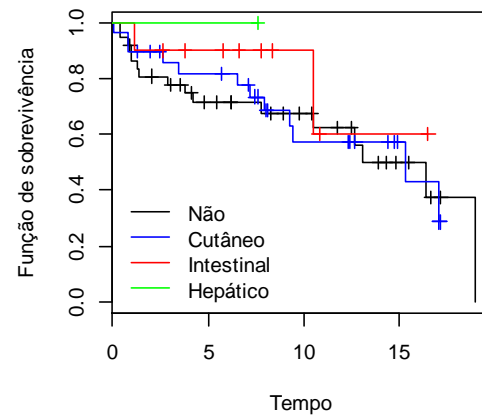
Tacrolimus



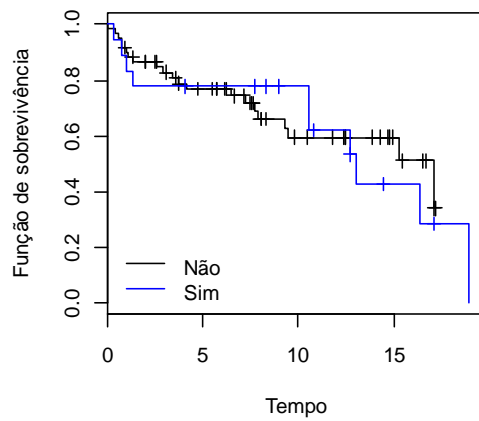
Corticoide



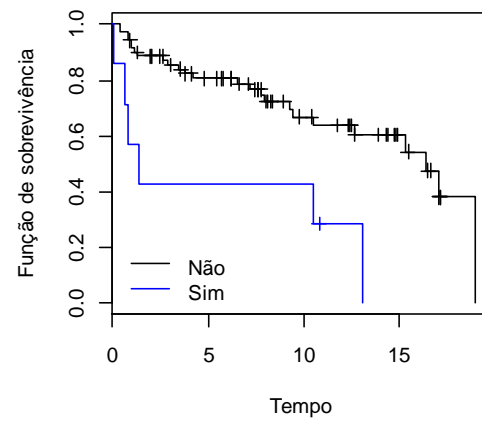
GVHD



Hipertensão



Dislipidemia



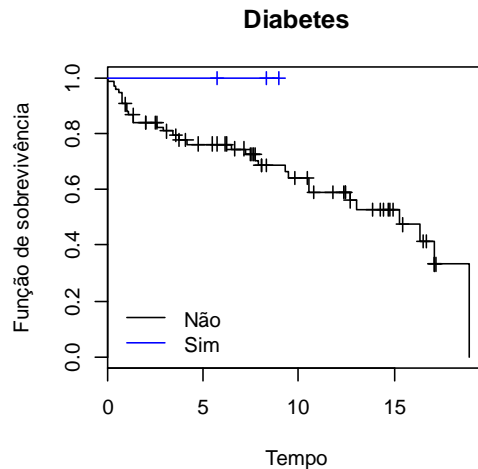


Figura 7: Estimativa de Kaplan-Meier da função de sobrevivência obtida a partir do estimador de para as variáveis: Diagnóstico, Género, Idade no transplante, IMC no transplante, TBI, Ciclosporina, Mofetil, Tacrolimus, Corticoide, GVHD, Hipertensão, Dislipidemia e Diabetes.

A estimativa da função de risco cumulativa foi obtida a partir do estimador não paramétrico de Kaplan-Meier e do estimador de Nelson-Aalen e, como seria de esperar, nos instantes iniciais as curvas das estimativas são sobreponíveis, e observa-se uma maior discrepância nos últimos instantes (figura 8).

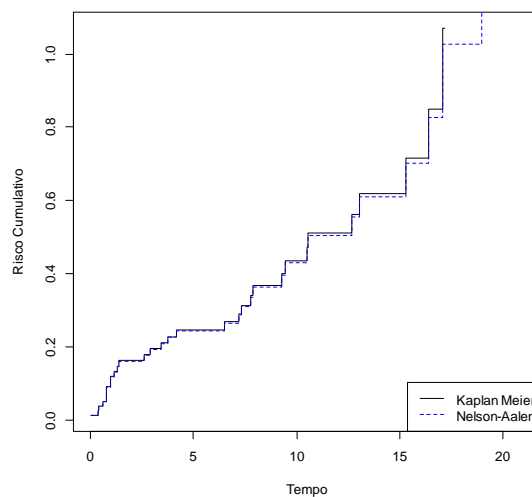


Figura 8: Comparação de estimativas da função de risco.

3.5 Comparação entre curvas de sobrevivência

Na comparação entre curvas de sobrevivência foram utilizados os testes não paramétricos log-rank, Gehan e Tarone-Ware (tabela 7). Foi necessário categorizar as variáveis “idade no transplante” e “IMC no transplante”, o que deu origem às variáveis

idade categorizada (<35, ≥35) e IMC categorizada (≤ 24.9 , >24.9). Nas diferentes análises efetuadas não se verificaram diferenças estatisticamente significativas, com a exceção da dislipidemia e IMC no transplante categorizada. Os indivíduos com dislipidemia apresentaram uma mediana de tempo até à ocorrência do acontecimento de interesse estimada de 1.38 anos, enquanto os doentes sem dislipidemia apresentaram uma mediana estimada de 16.38 anos. No caso dos doentes com obesidade ou excesso de peso no transplante, a estimativa da mediana do tempo até à ocorrência do acontecimento de interesse foi de 15.3 anos, ao passo que nos indivíduos com peso normal no transplante a mediana estimada foi de 10.53 anos.

Variáveis	Tipo de teste		
	Log Rank	Gehan	Tarone-Ware
Diagnóstico	0.228	0.266	0.246
Género	0.916	0.757	0.825
Idade Transplante Categorizada	0.197	0.288	0.232
IMC Transplante Categorizada	0.106	0.057	0.076
TBI	0.436	0.431	0.434
Ciclosporina	0.177	0.216	0.196
Mofetil	0.504	0.416	0.458
Tacrolimus	0.706	0.651	0.678
Corticoides	0.778	0.839	0.810
GVHD	0.797	0.770	0.783
HTA	0.794	0.868	0.834
Dislipidemia	0.002	0.0015	0.0017
Diabetes	0.313	0.315	0.313

Tabela 7: Testes não paramétricos log-rank, Gehan e Tarone-Ware. Os valores apresentados referem-se ao valor-p dos testes efetuados.

3.6 Modelo de Cox

As covariáveis que apresentaram um valor-p mais baixo nos testes não paramétricos ou que são consideradas pertinentes ponto de vista clínico, foram incluídas no modelo. É de referir que optou-se por não incluir a variável Dislipidemia, visto que para alguns doentes os dados relativos a esta variável foram colhidos durante o tempo de seguimento e não no início do estudo.

A permissa de proporcionalidade das funções de risco foi analisada com base nos gráficos $\log(-\log \hat{S}(t))$ contra o tempo, para as variáveis que serão incluídas no modelo (figura 9 e 10). Nos gráficos relativos à variável idade categorizada, IMC categorizada e TBI verifica-se que as curvas cruzam-se, o que pode indicar que o pressuposto de proporcionalidade das funções de risco não é válido. No caso da variável diagnóstico observa-se que as curvas são muito próximas, exceto no caso da categoria LMA. Portanto, parecem existir dúvidas quanto à proporcionalidade das funções de risco.

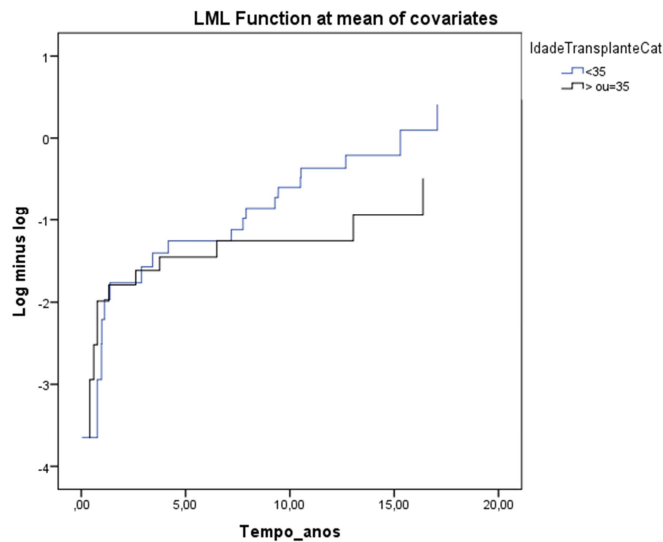


Figura 9: Gráficos $\log(-\log \hat{S}(t))$ em função do tempo para as variáveis Idade no transplante categorizada.

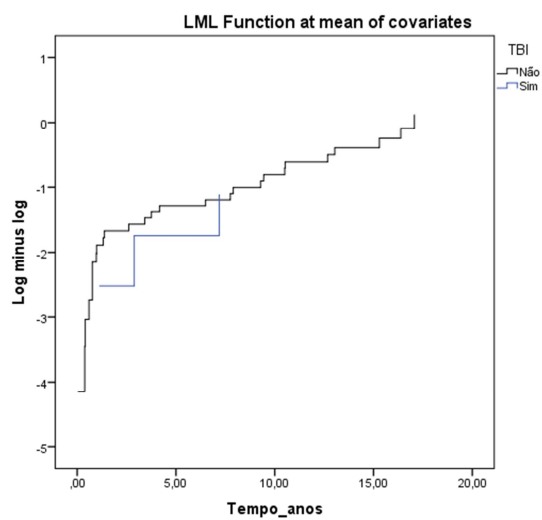
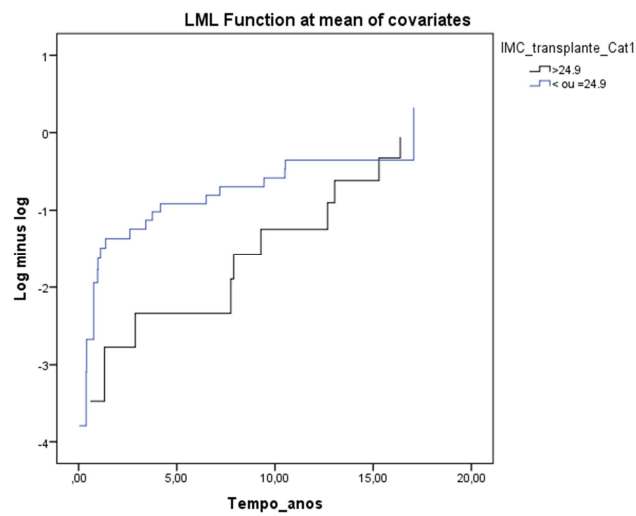
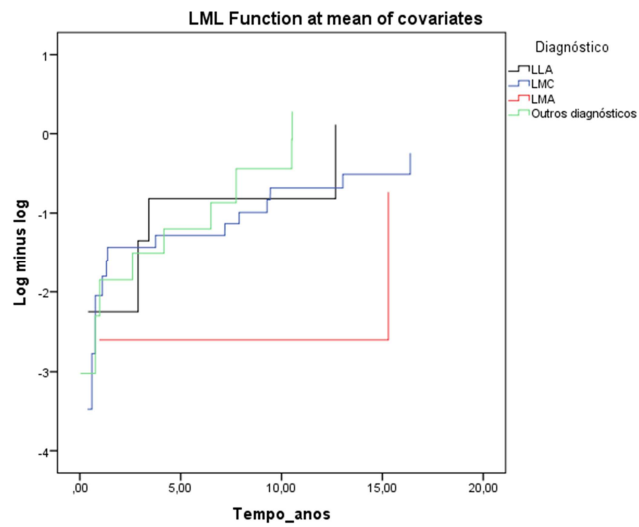


Figura 10: Gráficos $\log(-\log \hat{S}(t))$ em função do tempo para as variáveis Diagnóstico, IMC no transplante categorizada e TBI.

A metodologia de seleção de covariáveis utilizada foi descrita anteriormente na secção 2.4 do Capítulo 2.

Covariáveis:

x_1 -Idade no transplante categorizada

x_2 -Diagnóstico

x_3 -IMC no transplante categorizada

x_4 -TBI

	$-2 \log \hat{L}$	valor-p
Modelo nulo	111.21	
x_1	110.98	0.6315
x_2	109.01	0.1380
x_3	109.87	0.2470
x_4	110.88	0.5656
$x_2 + x_3$	108.03	0.2488
$x_2 + x_3 + x_1$	107.69	0.7641
$x_2 + x_3 + x_4$	106.70	0.2488

Tabela 8: Estatística $-2 \log \hat{L}$ e respetivo valor-p para cada modelo ajustado.

O modelo de Cox foi utilizado na análise dos fatores com influência no tempo decorrido entre a paragem da imunossupressão até à alteração do IMC para uma categoria superior correspondente a excesso de peso ou obesidade. Não tendo nenhuma das covariáveis revelado influência significativa no tempo, optou-se por incluir no modelo as covariáveis que mais fizeram diminuir a estatística $-2 \log \hat{L}$ (tabela 8). Ajustando aos dados este modelo de Cox, obteve-se a seguinte estimativa da função de risco:

$$\hat{h}(t; x) = \hat{h}_0(t) \exp(-0.548x_{IMCcat} - 0.366x_{LMC} - 1.066x_{LMA} + 0.0975x_{outros diag})$$

3.7 Análise de resíduos para o modelo de Cox

Resíduos de Cox-Snell

Através da análise dos resíduos de Cox-Snell é possível avaliar o ajustamento global do modelo. Na figura 11, observa-se a estimativa da função de risco cumulativa calculada com base no estimador de Nelson-Aalen *versus* os resíduos de Cox-Snell, e ajustando uma reta com ordenada na origem nula e declive um, conclui-se que o modelo parece ser adequado.

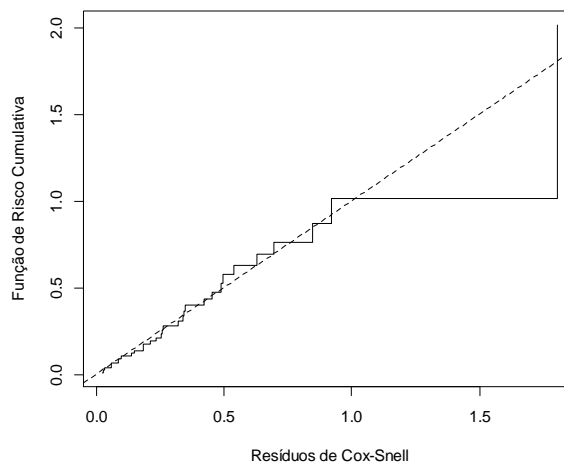


Figura 11: Estimativa da função de risco cumulativa para os resíduos *versus* os resíduos de Cox-Snell.

Resíduos de Schoenfeld

A proporcionalidade dos riscos foi avaliada com base na representação gráfica dos resíduos de Schoenfeld (figura 12). Através destes gráficos é razoável presumir-se a existência de proporcionalidade das funções de risco, visto que não é encontrada uma tendência marcada nos resíduos Schoenfeld em função do tempo. Esta interpretação é reforçada pelo teste estatístico que permite testar a existência de correlação entre os resíduos de Schoenfeld e o tempo. A tabela 9 apresenta os valores-p relativos a este teste, sendo possível observar que em nenhuma das variáveis é rejeitada a hipótese nula de que não existe correlação entre os resíduos e o tempo, assim sendo admite-se a existência de proporcionalidade dos riscos para todas as covariáveis, embora no caso

da variável IMC no transplante categorizada a análise do gráfico, bem como o valor- $p=0.0654$ obtido no teste coloquem algumas dúvidas.

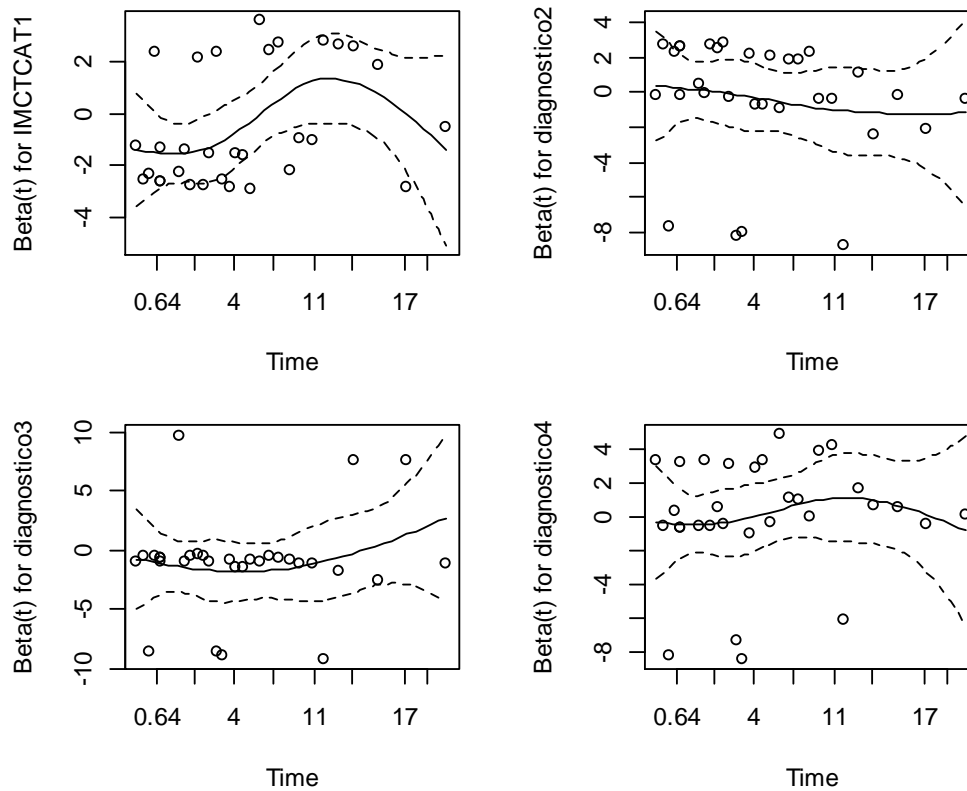


Figura 12: Análise da premissa de riscos proporcionais para a covariável diagnóstico e IMC categorizada com os resíduos de Schoenfeld padronizados *versus* tempo transformado.

	ρ	χ^2	valor-p
IMCcat	0.330	3.394	0.0654
DiagnosticoLMC	-0.155	0.796	0.3723
Diagnóstico LMA	0.152	0.640	0.4237
Diagnóstico Outros	0.100	0.310	0.5777
Global	NA	7.386	0.1168

Tabela 9: Resultados do teste de correlação entre os resíduos de Schoenfeld. Valor-p associado ao teste.

Resíduos Martingala

A figura 13 apresenta o gráfico dos resíduos Martingala do modelo ajustado versus o índice do indivíduo, com curva de suavização LOWESS (*Locally Weighted Scatterplot Smoother*). Através deste gráfico observa-se que não parece existir um padrão, o que indica um bom ajustamento do modelo aos dados.

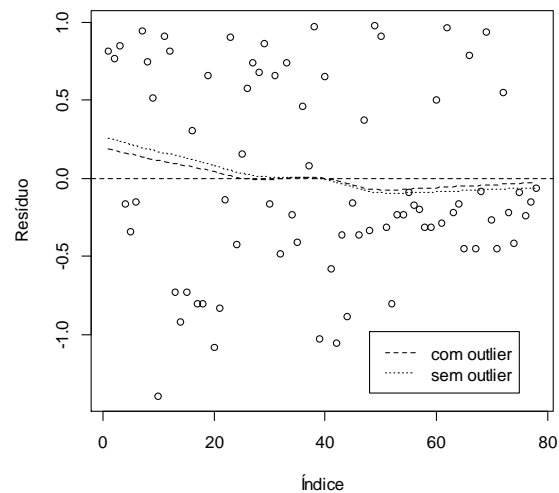


Figura 13:Resíduos Martingala do modelo ajustado versus índice do indivíduo.

Desvios residuais

As figuras 14 e 15 mostram os gráficos dos desvios residuais *versus* o índice do indivíduo e do tempo, respetivamente. Tendo em conta de que nestes gráficos não é observado um padrão, admite-se que o modelo é adequado.

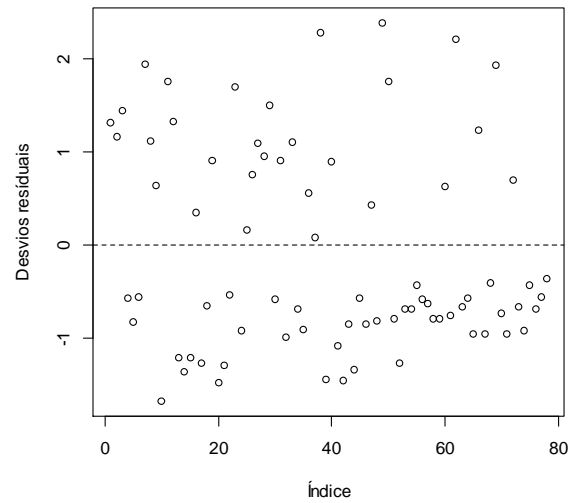


Figura 14: Desvios residuais *versus* índice.

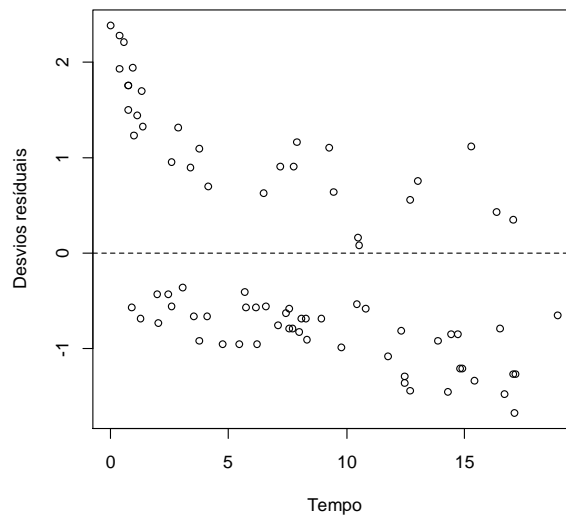


Figura 15: Desvios residuais *versus* o tempo.

Resíduos Score

Através da figura 16 é possível verificar que em ambas as covariáveis incluídas no modelo são encontrados valores extremos, o que pode indicar problemas no ajustamento do modelo aos dados.

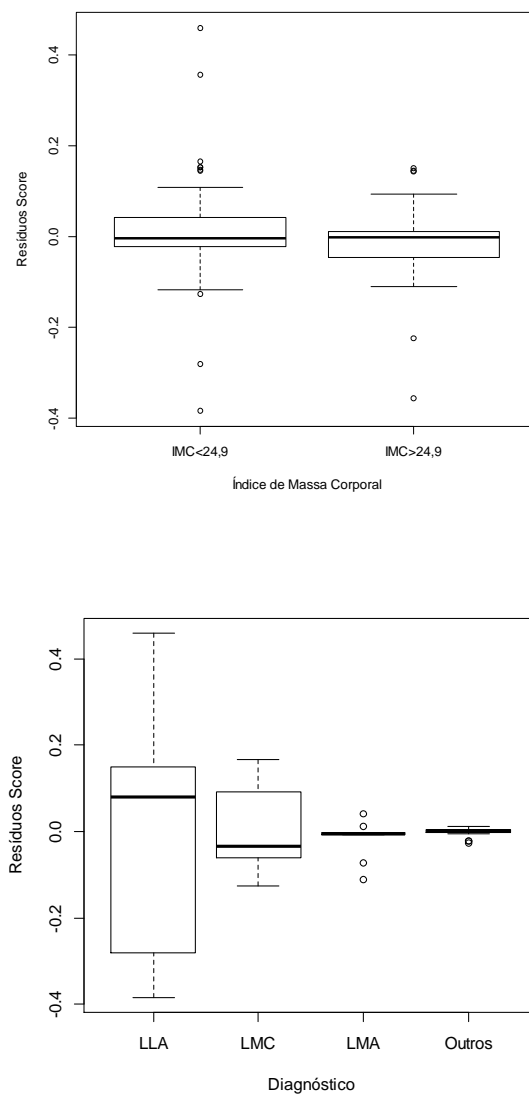


Figura 16: Resíduos score para as covariáveis IMC no transplante e Diagnóstico.

3.8 Modelos de regressão paramétricos

A tabela 10 mostra as estimativas dos parâmetros e os valores de AIC para os seguintes modelos de regressão paramétricos: exponencial, Weibull, log-logístico e log-normal.

	Exponencial			Weibull			Log logístico			Log normal		
	$\hat{\alpha}_j$	Erro Padrão	valor-p	$\hat{\alpha}_j$	Erro Padrão	valor-p	$\hat{\alpha}_j$	Erro Padrão	Valor-p	$\hat{\alpha}_j$	Erro Padrão	P value
Intercept	2.55	0.52	8.95x10 ⁻⁷	2.61	0.60	1.57x10 ⁻⁵	1.98	0.711	0.0053	2.04	0.76	0.0074
DiagLMC	0.22	0.56	0.69	0.21	0.65	0.74	0.22	0.75	0.76	0.13	0.80	0.86
DiagLMA	0.99	0.76	0.19	1.10	0.89	0.21	1.36	0.96	0.15	1.41	1.02	0.16
DiagOutros	-0.09	0.60	0.87	-0.08	0.69	0.90	0.15	0.81	0.84	0.03	0.87	0.97
IMCCAT	0.48	0.38	0.21	0.56	0.45	0.21	0.92	0.51	0.06	1.10	0.55	0.04
Escala	1			1.15	0.14	0.36	0.02	0.13	0.91	0.63	0.13	1.78x10 ⁻⁶
AIC		251.63			252.78			253.26			254.07	

Tabela 10: Estimativas dos parâmetros e os valores de AIC para os modelos de regressão exponencial, Weibull, log-logístico e log-normal

De acordo com os valores obtidos, optou-se por continuar a análise estatística com os modelos exponencial e Weibull visto apresentarem o valor de AIC mais baixo.

Modelo de regressão exponencial

O modelo exponencial é concomitantemente um modelo de riscos proporcionais e de tempo de vida acelerado. Assim, a expressão na forma de modelo de tempo de vida acelerado é dada por:

$$\log T = \mu + \alpha'x + \sigma\varepsilon.$$

O modelo estimado é então:

$$\log T = 2,55 + 0,22z_{LMC} + 0,99z_{LMA} - 0,99 z_{Outros diag} + 0,48z_{IMCcat}) + \varepsilon.$$

A função de sobrevivência, dado o vetor de covariáveis x , é dada por:

$$S(t; x) = \exp[-\exp(-\mu)t \exp((- \alpha'x)].$$

A função de sobrevivência do modelo exponencial na forma de modelo de riscos proporcionais é dada por

$$S(t|x) = \exp(-\lambda t \exp(\beta'x))$$

portanto, tem-se que

$$\lambda = \exp(-\mu) \text{ e } \beta_j = -\alpha_j.$$

Como a função de risco do modelo de riscos proporcionais é dada por

$$h(t|x) = \lambda \exp(\beta'x)$$

a sua estimativa é:

$$\hat{h}(t|x) = 0.078 \exp(-0.22x_{LMC} - 0.99x_{LMA} + 0.09x_{Outros diag} - 0.48x_{IMCcat}).$$

Modelo de regressão Weibull

Em primeiro lugar, considerou-se o modelo Weibull expresso como modelo de tempo de vida acelerado, visto ser a forma disponível no *software* estatístico. No entanto, através das relações que se podem estabelecer entre os parâmetros para o modelo Weibull na forma log-linear e os parâmetros no modelo de riscos proporcionais, foi também possível uma representação como modelo de riscos proporcionais. Assim, no caso do modelo de tempo de vida acelerado temos que:

$$\log T = \mu + \alpha'x + \sigma\varepsilon.$$

O modelo estimado é:

$$\log T = 2.61 + 0.21x_{LMC} + 1.1x_{LMA} - 0.085 x_{Outros diag} + 0.56x_{IMCcat} + 1.15\varepsilon$$

A função de sobrevivência dado o vetor de covariáveis x é dada por:

$$S(t; x) = \exp[-\exp(-\mu/\sigma)t^{1/\sigma} \exp((-\alpha/\sigma)'x)]$$

Comparando a função de sobrevivência para o modelo de tempo de vida acelerado, com a função de sobrevivência do modelo de riscos proporcionais, que é dada por

$$S(t; x) = \exp(-\lambda t^\gamma \exp(\beta'x)),$$

podemos estabelecer as seguintes relações:

$$\lambda = \exp(-\mu/\sigma), \gamma = 1/\sigma \text{ e } \beta_j = -\alpha_j/\sigma.$$

Como modelo de riscos proporcionais, a função de risco é dada por

$$h(t; x) = h_0(t) \exp(\beta' x) = \lambda \gamma t^{\gamma-1} \exp(\beta' x),$$

sendo $\lambda \exp(\beta' x)$ o parâmetro de escala e γ o parâmetro de forma. Neste caso, no instante t , a estimativa do risco de subir na categoria do IMC, tendo em conta a influência das covariáveis diagnóstico e IMC no transplante e de acordo com o modelo Weibull, é dada por:

$$\hat{h}(t|x) = 0.103 * 0.869t^{0.869-1} \exp(-0.18x_{LMC} - 0.95x_{LMA} + 0.073x_{Outros\ diag} - 0.486 x_{IMCcat})$$

$$\hat{h}(t|x) = 0.089t^{-0.131} \exp(-0.18x_{LMC} - 0.95x_{LMA} + 0.073x_{Outros\ diag} - 0.486 x_{IMCcat})$$

3.9 Análise de resíduos para os modelos paramétricos

Modelo exponencial

Resíduos de Cox-Snell

Os resíduos foram definidos considerando o modelo exponencial, assim uma estimativa da função de risco cumulativa para o indivíduo padrão é $\hat{H}_0(t_i) = -\ln(S(t)) = \alpha t_i$, logo $\hat{r}_i = \alpha \exp(\beta' x_i) t_i$. A figura 17 mostra os resíduos de Cox-Snell vs a estimativa da função de risco cumulativa para os resíduos calculada com base no estimador de Nelson-Aalen, sendo possível observar que o modelo parece ser adequado.

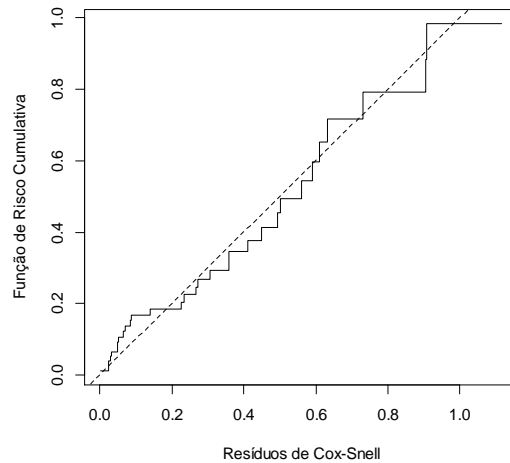


Figura 17: Estimativa da função de risco cumulativa para os resíduos *versus* os resíduos de Cox-Snell.

Resíduos Martingala

A figura 18 mostra os resíduos Martingala *versus* o índice do indivíduo sendo possível observar os valores de M_i que são superiores ou inferiores aos valores observados e que por sua vez indicam uma subestimação ou uma sobrestimação no número de acontecimentos de interesse. Assim sendo, este gráfico permite identificar indivíduos mal ajustados ao modelo. Nesta análise, o gráfico obtido é semelhante ao que foi obtido para o modelo de Cox, sendo que não parece existir um padrão, o que indica um bom ajustamento do modelo aos dados.

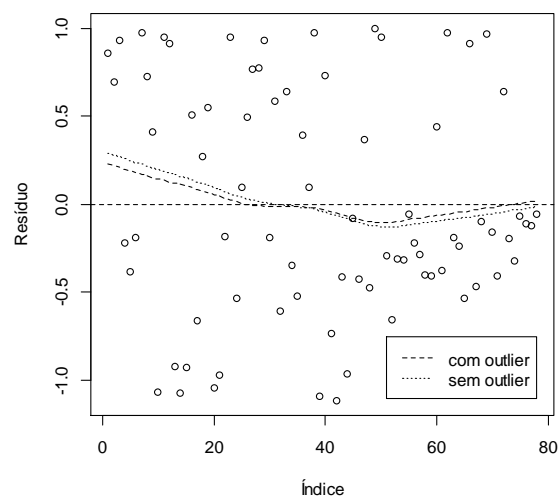


Figura 18: Resíduos Martingala do modelo ajustado *versus* índice do indivíduo.

Desvios residuais

A figura 19 e 20 mostra os desvios residuais versus o índice do indivíduo e os desvios residuais versus o tempo, respetivamente.

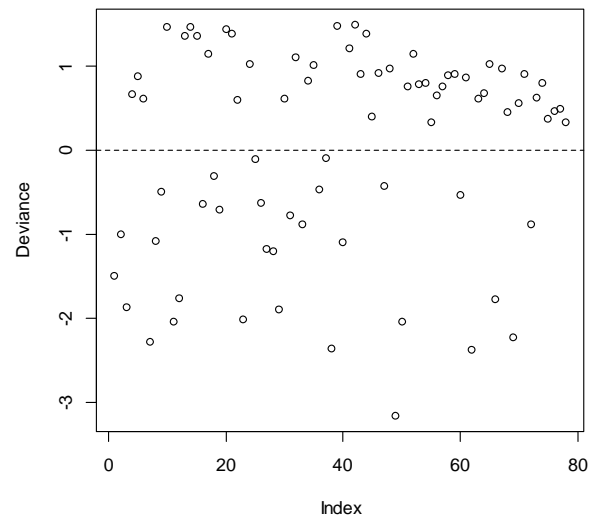


Figura 19: Desvios residuais *versus* índice.

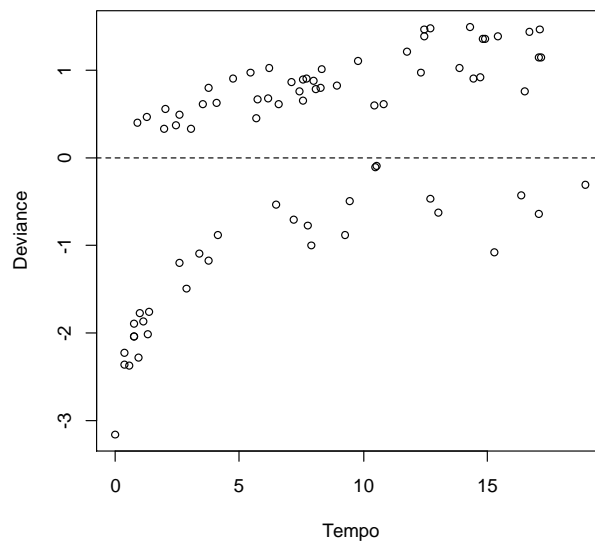


Figura 20: Desvios residuais *versus* tempo.

Modelo Weibull

A análise de resíduos foi igualmente efetuada para o modelo Weibull, e foram obtidos gráficos muito semelhantes ao modelo exponencial.

Resíduos de Cox-Snell

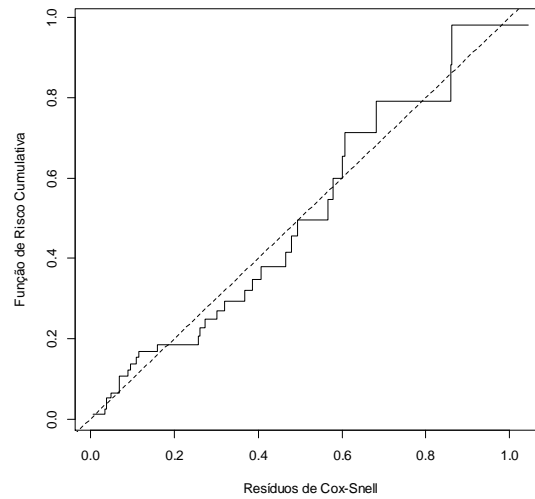


Figura 21: Estimativa da função de risco cumulativa *versus* os resíduos de Cox-Snell.

Resíduos Martingala

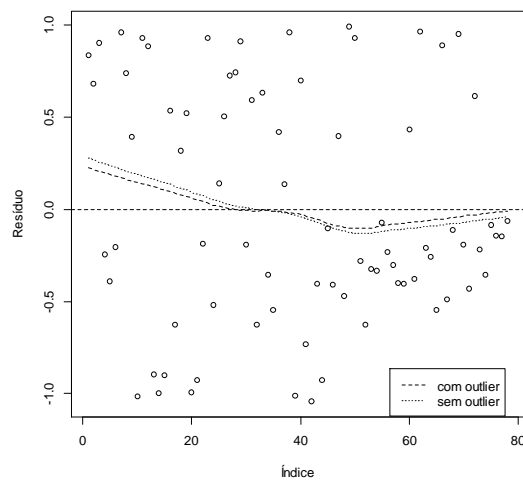


Figura 22: Resíduos Martingala do modelo ajustado *versus* índice do indivíduo.

Desvios residuais

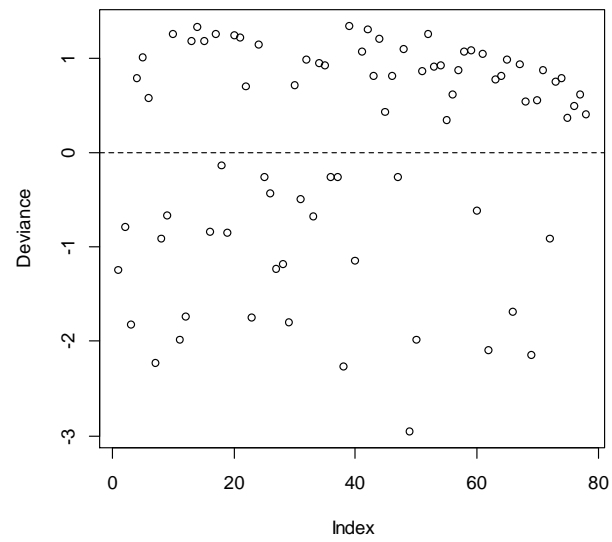


Figura 23: Desvios residuais *versus* índice.

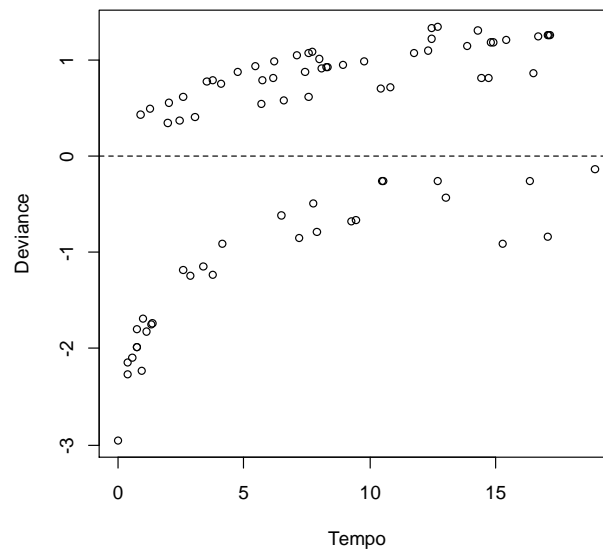


Figura 24: Desvios residuais *versus* índice.

3.10 Comparação gráfica de modelos

Foi efetuada a comparação da estimativa da função de sobrevivência obtida pelo estimador não paramétrico de Kaplan-Meier com as estimativas obtidas com base nos modelos paramétricos exponencial e Weibull. A figura 25 mostra o gráfico com as curvas de sobrevivência obtidas, sendo possível observar que existe algum afastamento da curva de Kaplan-Meier relativamente às estimativas paramétricas da função de sobrevivência.

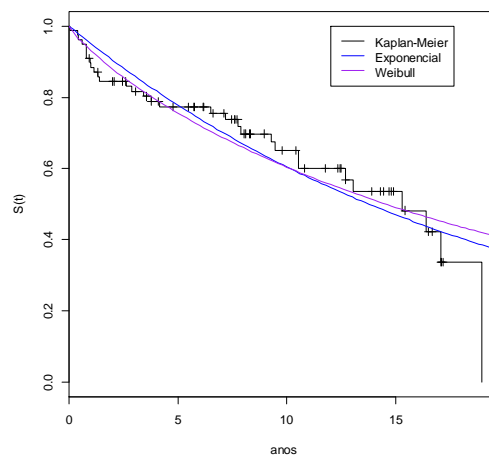


Figura 25: Estimativas da função de sobrevivência obtidas com modelos paramétricos (exponencial e Weibull) e com o estimador de Kaplan-Meier.

Capítulo 4-Resultados

4.1 Modelo de Cox

Covariáveis	$\hat{\beta}_j$	Erro Padrão	$\exp(\hat{\beta}_j)$	IC 95%	valor-p
IMCTransplante \leq 24.9	-	-	1.00	-	-
IMCTransplante $>$ 24.9	-0.548	0.401	0.577	[0.26-1.27]	0.172
LLA	-	-	1.00	-	-
LMC	-0.366	0.581	0.693	[0.22-2.16]	0.528
LMA	-1.066	0.772	0.344	[0.07-1.56]	0.167
Outros diagnósticos	0.0975	0.609	1.102	[0.33-3.64]	0.873

Tabela 11: Estimativas dos parâmetros e intervalos de 95% de confiança obtidos a partir do modelo de Cox.

De acordo com a tabela 11 verifica-se que não existe evidência de associação entre as covariáveis estudadas e o risco de ocorrer um incremento no IMC que permita a passagem do indivíduo para uma categoria superior de excesso de peso ou obesidade. Efetivamente, a última coluna mostra os valores-p obtidos pelo teste de Wald, que testa a hipótese nula de $\beta_j = 0$, e para nenhuma das covariáveis é rejeitada esta hipótese.

Para interpretarmos os valores obtidos é importante ter em conta que coeficientes positivos indicam um aumento no risco de subir uma categoria no IMC e coeficientes negativos indicam covariáveis que contribuem para a redução desse risco. Assim, estima-se que um indivíduo com IMC na data do transplante superior a 24.9 tem uma redução de 42.3% no risco de aumentar uma categoria no IMC, relativamente a um indivíduo com IMC inferior ou igual a 24.9 e com o mesmo diagnóstico. No que diz respeito ao diagnóstico, estima-se nos indivíduos com LMC e LMA uma redução no risco de subir uma categoria no IMC de 30.7% e 65.5%, respetivamente, em relação a um indivíduo com LLA e na mesma categoria de IMC no transplante. Por último, um indivíduo classificado com outros diagnósticos tem um aumento estimado de 10.2% no risco de aumentar uma categoria no IMC, comparativamente a um indivíduo com LLA e na mesma categoria do IMC no transplante.

No entanto, tendo em conta que a análise de resíduos sugeriu a não proporcionalidade dos riscos para a covariável IMC no transplante, foi igualmente ajustado um modelo de Cox com inclusão da covariável diagnóstico, mas estratificado na covariável IMC no transplante. A tabela 12 mostra os resultados obtidos nesta análise.

Covariáveis	$\hat{\beta}_j$	Erro padrão	$\exp(\hat{\beta}_j)$	valor-p
LLA	-	-	1.00	-
LMC	-0.4127	0.662	0.586	0.48
LMA	-1.1786	0.308	0.781	0.13
Outros diagnósticos	0.0171	1.017	0.614	0.98

Tabela 12: Estimativas dos parâmetros obtidos a partir do modelo de Cox estratificado para o IMC no transplante e com inclusão da covariável diagnóstico.

Na análise estratificada pode-se constatar que não se verificaram alterações muito significativas nos valores obtidos, sendo no entanto de salientar que o valor-p correspondente ao diagnóstico de LMA ficou mais próximo da significância estatística ($p=0.13$). Assim sendo, estima-se para um indivíduo com o diagnóstico de LMA, uma redução de aproximadamente 22% no risco de uma subida na categoria de IMC, com alteração para uma categoria de excesso de peso ou obesidade, relativamente a um indivíduo com LLA.

4.2 Modelos de regressão paramétricos

Apesar do modelo exponencial ter um valor de AIC inferior optou-se por indicar os resultados do modelo Weibull, visto que não parece ser adequado assumir que o risco de aumento no IMC é constante ao longo do tempo, sendo por isso mais razoável considerar um modelo mais flexível.

Covariável	$\hat{\alpha}_j$	Erro padrão	$\exp(\hat{\alpha}_j)$	Erro padrão
IMCTransplante \leq 24.9	-	-	1.00	-
IMCTransplante $>$ 24.9	0.56	0.452	1.57	0.172
LLA	-	-	1.00	-
LMC	0.211	0.650	1.23	0.528
LMA	1.108	0.891	3.02	0.167
Outros diagnósticos	-0.085	0.697	0.91	0.873

Tabela 13: Estimativas dos parâmetros obtidos a partir do modelo Weibull na forma de modelo de tempo de vida acelerado.

A tabela 13 apresenta as estimativas dos parâmetros obtidas a partir do modelo Weibull na forma de modelo de tempo de vida acelerado. É de notar que tal como no caso do modelo de Cox, não foi verificada uma influência estatisticamente significativa das covariáveis incluídas no modelo no tempo até à subida do IMC.

Sabe-se que

$$T_x = T_0 \exp(\alpha'x)$$

em que T_x é o tempo de vida para um indivíduo com o vetor de covariáveis x e T_0 o tempo de vida para o indivíduo padrão. Logo, um coeficiente positivo indica que um indivíduo tem um tempo até à subida na categoria de IMC que é mais longo do que o indivíduo padrão. Em alternativa, um coeficiente negativo indica uma aceleração no tempo até à subida de uma categoria do IMC, relativamente ao indivíduo padrão.

Assim sendo, no que diz respeito ao IMC no transplante observa-se que um indivíduo que apresenta excesso de peso ou obesidade (IMC superior a 24.9) demora mais tempo a subir uma categoria no IMC, em relação ao indivíduo com IMC inferior ou igual a 24.9 e com o mesmo diagnóstico. Do mesmo modo, um indivíduo com diagnóstico de LMC ou LMA leva mais tempo até à subida de uma categoria no IMC, relativamente a um indivíduo com diagnóstico de LLA e na mesma categoria do IMC no transplante. Por último, um indivíduo incluído na categoria de outros diagnósticos apresenta uma aceleração no tempo até à subida na categoria de IMC em relação a um indivíduo com diagnóstico de LLA e na mesma categoria do IMC no transplante.

Covariável	Modelo Weibull		Modelo de Cox	
	$\hat{\beta}_j$	$\exp(\hat{\beta}_j)$	$\hat{\beta}_j$	$\exp(\hat{\beta}_j)$
IMCTransplante≤24.9	-	1.00	-	1.00
IMCTransplante>24.9	-0.486	0.615	-0.548	0.577
LLA	-	1.00	-	1.00
LMC	-0.18	0.835	-0.366	0.693
LMA	-0.95	0.386	-1.066	0.344
Outros diagnósticos	0.073	1.075	0.0975	1.102

Tabela 14: Estimativas dos parâmetros obtidas a partir do modelo Weibull na forma de modelo de riscos proporcionais e para o modelo de Cox.

A tabela 14 apresenta as estimativas dos parâmetros obtidas a partir do modelo Weibull na forma de modelo de riscos proporcionais e também para as estimativas obtidas para o modelo de Cox. Através dos valores obtidos, estima-se que um doente com IMC correspondente a excesso de peso ou obesidade na data do transplante tem uma redução de 38.5% do risco de aumentar uma categoria no IMC, relativamente a um indivíduo com IMC correspondente a peso normal (IMCTransplante≤24.9) e com o

mesmo diagnóstico. Da mesma maneira, estima-se que indivíduos com diagnóstico de LMC e LMA têm uma redução no risco de aumentar uma categoria no IMC de 16.5% e 61.4%, respectivamente, relativamente a um indivíduo com diagnóstico de LLA e na mesma categoria de IMC no transplante. Por último, um indivíduo com outros diagnósticos tem um risco 7.5% superior de subida na categoria de IMC, comparativamente a um indivíduo com LLA e na mesma categoria de IMC no transplante. Como seria de esperar, os resultados obtidos para o modelo Weibull são semelhantes aos que foram obtidos para o modelo de Cox.

4.3 Discussão e Conclusão

Este estudo teve como objetivo analisar o tempo até à alteração do IMC para uma categoria superior correspondente a excesso de peso ou obesidade após TCPH-alo. Assim sendo, verificou-se que a estimativa da mediana do tempo até ao acontecimento de interesse com base no estimador de Kaplan-Meier foi de 15.3 anos. Atendendo a que a idade média dos indivíduos na última observação foi de 44.6 anos [22.43-70.47], podemos verificar que para esta classe etária a prevalência de excesso de peso obtida foi inferior ao valor obtido para a população portuguesa saudável (48.1% vs. 43.9%), no entanto a prevalência de obesidade foi semelhante em ambos os estudos (17.8% vs 18.3%) (Carmo, et al., 2008). Este resultado sugere que o TCPH-alo não parece estar implicado no desenvolvimento de excesso de peso ou obesidade, mas ainda assim é de notar que a prevalência obtida é elevada e poderá colocar este grupo de doentes em risco de outras doenças.

Num estudo recente (Lynce, et al., 2012) que incluiu 219 doentes com diagnóstico de linfoma seguidos retrospectivamente por um período de 18 meses, observou-se que cerca de 9% dos indivíduos tiveram um aumento de peso superior a 20% durante o período de seguimento. Neste estudo constatou-se ainda que no diagnóstico 34.7% dos indivíduos apresentava excesso de peso e 34.3% obesidade; em contrapartida, aos 18 meses a prevalência de excesso de peso e de obesidade era de 37.3% e 35.9%, respetivamente. Estes resultados são concordantes com os resultados obtidos no presente estudo, visto ter sido observado um aumento na percentagem de indivíduos com excesso de peso ou obesidade ao longo do estudo (excesso de peso-inicial:18.3% e final: 43.9%; obesidade inicial: 7.3% e final 18.3%). No entanto, verificou-se no estudo anteriormente citado uma prevalência de obesidade muito superior, que pode ser explicada pelo facto de ter sido realizado nos Estados Unidos da América, onde a prevalência de obesidade é superior à que é encontrada em Portugal (35.1% vs. 14.4%).

Na análise multivariada, com recurso a um modelo de regressão semi-paramétrico (modelo de Cox) e a um modelo paramétrico (modelo Weibull) as covariáveis estudadas não revelaram influência estatisticamente significativa no tempo até à subida do IMC com passagem do indivíduo para uma categoria superior do IMC. No entanto, é importante salientar os resultados obtidos para as covariáveis diagnóstico e IMC no transplante. No que diz respeito ao diagnóstico, estima-se nos indivíduos com LMC e LMA uma redução no risco de subir uma categoria no IMC de 30.7% e 65.5%, respetivamente, em relação a um indivíduo com LLA e na mesma categoria de IMC no transplante. Para além disto, um indivíduo classificado com outros diagnósticos tem um aumento estimado de 10,2% no risco de aumentar uma categoria no IMC, comparativamente a um indivíduo com LLA e na mesma categoria do IMC no transplante. Contudo, não foi encontrada evidência científica que fundamentasse estes resultados, pelo que deverão ser confirmados. No caso do IMC no transplante, estima-se que um indivíduo com IMC na data do transplante superior a 24.9 tem uma redução de 42.3% no risco de aumentar uma categoria no IMC, relativamente a um indivíduo com IMC inferior ou igual a 24.9 e com o mesmo diagnóstico. Assim sendo, parecem ser os indivíduos com peso normal no transplante que estão em maior risco de subir na categoria do IMC, ou seja de virem a apresentar excesso de peso ou obesidade passados alguns anos. Atendendo a que não foi encontrada uma associação significativa com o tipo de terapêutica, este resultado sugere que poderá existir uma alteração no estilo de vida deste grupo particular de doentes, nomeadamente na atividade física e no padrão alimentar, que não é tão pronunciada nos doentes que apresentam excesso de peso ou obesidade no transplante. Efetivamente, a literatura mostra que o aumento de peso nos sobreviventes de cancro pode ter a ver com alterações no estilo de vida no contexto do tratamento antineoplásico, devido a complicações somáticas ou a fatores psicológicos (Haas, et al., 2010). Nos sobreviventes de cancro infantil e de cancro da mama, verificou-se uma menor atividade física, que poderá explicar o aumento ponderal (Haas, et al., 2010). Num estudo prospetivo realizado em 120877 indivíduos foi obtida uma taxa de aumento de peso de 1.5kg por cada 4 anos, sendo os principais alimentos implicados as batatas fritas, bebidas com açúcar, carnes vermelhas processadas e não processadas. Em contrapartida, foi obtida uma associação inversa entre o peso e a ingestão de legumes, cereais integrais, frutas, frutos oleaginosos e iogurte. Para além disto, outros fatores associados ao estilo de vida nomeadamente o consumo de bebidas alcoólicas, tabagismo, atividade física, número de horas de sono e passadas a ver televisão, apresentaram uma associação independente com a alteração do peso (Mozaffarian, et al., 2011).

Neste estudo não foi encontrada evidência de associação entre o tratamento farmacológico e o aparecimento de excesso de peso/obesidade, inclusivamente não foi encontrada uma diferença estatisticamente significativa entre a média de peso após TCPH-alo e na paragem da imunossupressão, sendo de 65.04kg e 65.63kg

($p=0.453$), respetivamente. Na literatura, o aumento de peso durante os tratamentos antineoplásicos tem sido atribuída à quimioterapia (sobretudo em doentes com cancro da mama), irradiação corporal total e corticoesteroides, embora sem ser de forma consistente. Serão necessários mais estudos para esclarecer esta associação. Contudo, de acordo com os dados obtidos o TCPH-alo não parece estar associado à etiologia do excesso de peso/obesidade, pelo que deverá ser dada maior atenção aos fatores inerentes ao estilo de vida, visto serem potencialmente modificáveis. De facto, uma das limitações do presente estudo tem a ver com não ter sido possível recolher informação de covariáveis que dizem respeito ao estilo de vida. Assim sendo, fica por esclarecer o efeito dessas covariáveis no desenvolvimento de obesidade/excesso de peso nos doentes sujeitos a TCPH-alo, sendo por isso necessários mais estudos para esclarecer esta questão.

Por último, é importante referir que foi possível constatar que efetivamente uma elevada percentagem dos doentes submetidos a TCPH-alo aumentam de peso embora sem influência aparente da terapêutica instituída. No entanto, seria importante continuar a estudar esta população, sendo de salientar que o próximo passo seria a estruturação de um estudo prospetivo, no qual fosse incluída uma análise mais detalhada da composição corporal com recurso a bioimpedância eléctrica, Tomografia Axial Computorizada e dinamometria, análises laboratoriais (incluindo endocrinologia) e variáveis relativas ao estilo de vida e estado psicológico.

Bibliografia

- Bastos, J., Rocha, C. (2006). Análise de Sobrevivência, Conceitos Básicos. *Arqui Med*, 20(5/6), 185-87.
- Bastos, J., Rocha, C. (2007). Análise de Sobrevivência, Métodos Não Paramétricos. *Arqui Med*, 4(3/4), 111-114.
- Carmo, I., Santos, O., Camolas, J., Vieira, J., Carreira, M., L. L. M., et al. (2008). Overweight and Obesity in Portugal: National Prevalence in 2003–2005. *Obesity Reviews*, 9, 11-19.
- Carvalho, M., Andreozzi, V., Codeço, C., Barbosa, M., & Shimakura, S. (2005). *Análise de Sobrevida*. Rio de Janeiro: Fiocruz.
- Ceschi, M., Gutzwiller, F., Moch, H., Eichholzer, M., & Probst-Hensch, N. M. (2007). Epidemiology and Pathophysiology of Obesity as a cause of Cancer. *Swiss Med Wkly*, 137, 50-56.
- Collett, D. (2003). *Modelling Survival Data in Medical Research*. Boca Raton: Chapman & Hall/CRC.
- Correia, A. (2007). *Rendibilidade de Transacções de Crédito Pessoal com Recurso a Análise de Sobrevivência*. Tese de Mestrado, Faculdade de Ciências da Universidade do Porto, Porto.
- Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society*, 34(2), 187-220.
- Dias, M. (2004). *Modelos de Tempo de Vida Acelerado em Análise de Sobrevivência*. Dissertação, Faculdade de Ciências da Universidade de Lisboa, Departamento de Estatística e Investigação Operacional, Lisboa.
- Diez, D. M. (2012). *Survival Analysis in R*. Obtido a 26 de Janeiro de 2013, de www.openintro.org/stat/surv.php
- Fuji, S., Kim, S.-W., Yoshimura, K.-i., Akiyama, H., Okamoto, S.-i. (2009). Possible Association Between Obesity and Posttransplantation Complications Including Infectious Diseases and Acute Graft-versus-Host Disease. *Biol Blood Marrow Transplant*, 15, 73-82.
- Haas, E. C., Oosting, S. F., Lefrandt, J. D., Enbattel, B. H., Sleijfer, D., Gietema, J. A. (2010). The Metabolic Syndrome in Cancer Survivors. *Lancet Oncol*, 193-203.
- Kaplan, E. L., Meier, P. (Junho de 1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282), 457-81.
- Kyle, U., Chalandon, Y., Mirabel, R., Karsegard, V., Hans, D., Trombetti, A. (2005). Longitudinal Follow-up of Body composition in Hematopoietic Stem Cell Transplant Patients. *Bone Marrow Transplantation*, 35, 1171-1177.

- Leucemia, A. P. (s.d.). Obtido a 17 de Janeiro de 2015, de <http://www.apcl.pt/dadores/o-que-e-a-medula-ossea>
- Lymphomation.org*. (s.d.). Obtido de <http://www.lymphomation.org/bmt-allo.htm>.
- Lynce, F., Pehlivanova, M., Catlett, J., & Alkovska, V. M. (Abril de 2012). Obesity in Adult Lymphoma Survivors. *Leukemia & Lymphoma*, 53(4), 569–57.
- Mahan, L. K., & Escott-Stump, S. (2010). *Alimentos, Nutrição e Dietoterapia*. Rio de Janeiro: Futura.
- Majhail, N., Flowers, M., Ness, K., Jagasia, M., Carpenter, P., Arora, M. (2009). High Prevalence of Metabolic Syndrome After Allogeneic Hematopoietic Stem Cell Transplantation. *Bone Marrow Transplant*, 43(1).
- Mantel, N., Haenszel, W. (Abril de 1959). Statistical Aspects of Analysis of Data From Retrospective Studies of Disease. *Journal of the National Cancer Institute*, 22(4), 719-49.
- Mozaffarian, D., Hao, T., Rimm, E. B., Willett, W. C., Hu, F. B. (2011). Changes in Diet and Lifestyle and Long-Term Weight Gain in Women and Men. *N Engl J Med*, 364, 2392-404.
- Muscaritoli, M., Grieco, G., Capria, S., Iori, A. P., Fanelli, F. R. (2002). Nutritional and Metabolic Support in Patients Undergoing Bone Marrow Transplant. *Am J Clin Nutr*, 75, 183-90.
- Rocha, C., Papoila, A. (2009). *Análise de Sobrevivência*. Lisboa: Instituto Nacional de Estatística.
- Rodríguez, G. (2010). *Parametric Survival Models*. Obtido a 5 de Abril de 2014, de <http://data.princeton.edu/pop509/ParametricSurvival.pdf>.
- Siviero-Miachon, A. A., Maria, A., Spinola-Castro, & Guerra-Junior, G. (2009). Adiposity in Childhood Cancer Survivors: Insights Into Obesity Physiopathology. *Arq Bras Endocrinol Metab*, 53(2), 190-200.
- Skibola, C. F. (2007). Obesity, Diet and Risk of Non-Hodgkin Lymphoma. *Cancer Epidemiol Biomarkers Prev*, 16(3), 392.
- Strom, S. S., Yamamura, Y., Kantarjian, H. M., & Cortes-Franco, J. E. (2009). Obesity, Weight Gain, and Risk of Chronic Myeloid Leukemia. *Cancer Epidemiol Biomarkers Prev*, 18, 1501-1506.
- Tan, P. H. (2004). Stem Cells: Current Usage and Future Potentials. *Annals Academy of Medicine*, 33:5.
- Therneau, T. (2013). *Package 'survival'*. Obtido a 3 de Março de 2013, de <http://r-forge.r-project.org>
- WHO, W. H. (s.d.). http://apps.who.int/bmi/index.jsp?introPage=intro_3.html. Obtido em 2013

Anexo I
Folha de Registo de dados

Nome: _____

Nº Processo: _____

Sexo: ____

Idade: _____

Data de Nascimento: __/__/____

Dados Clínicos	
Diagnóstico	
Terapêutica Farmacológica	Terapêutica pré transplante
	Quimioterapia (QT) Sim <input type="checkbox"/> Não <input type="checkbox"/> Tipo: _____
	Condicionamento
	Total Body Irradiation (TBI) Sim <input type="checkbox"/> Não <input type="checkbox"/> Dose: ____ Quimioterapia (QT) Tipo: _____
	Terapêutica pós transplante
	Medicação imunossupressora <ul style="list-style-type: none"> • ciclosporina <input type="checkbox"/> • tacrolimus <input type="checkbox"/> • corticóides <input type="checkbox"/> • Micofenolato de Mofetil <input type="checkbox"/> • Outro _____
Período de tempo entre o transplante e o presente	Data do dia de internamento: Data do transplante: Data da alta:
Tipo/nº de transplantes realizados	

Comorbilidades	DM	<input type="checkbox"/>	Data de aparecimento: __/__/__ Manteve-se até:
	HTA	<input type="checkbox"/>	Data de aparecimento: __/__/__ Manteve-se até:
	Dislipidemia	<input type="checkbox"/>	Data de aparecimento: __/__/__ Manteve-se até:
	GVHD	<input type="checkbox"/>	Data de aparecimento: __/__/__ Manteve-se até:

Dados Antropométricos	
Peso actual (kg)	
Peso à data de admissão (kg)	
Peso à data de alta (kg)	
Peso (kg) e data quando cessa medicação imunossupressora	
Altura (m)	
Data pós transplante de IMC > 24,9kg/m ²	
Última data no processo	

Anexo II

Estimador de Kaplan Meier e respectivos intervalos de confiança

a) Intervalos de confiança calculados com base no estimador de Kaplan Meier

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
0.030	78	1	0.987	0.0127	0.962	1.000
0.381	77	1	0.974	0.0179	0.939	1.000
0.410	76	1	0.962	0.0218	0.919	1.000
0.600	75	1	0.949	0.0250	0.900	0.998
0.770	74	3	0.910	0.0324	0.847	0.974
0.970	70	1	0.897	0.0344	0.830	0.965
0.980	69	1	0.884	0.0363	0.813	0.955
1.120	68	1	0.871	0.0380	0.797	0.946
1.320	66	1	0.858	0.0397	0.780	0.936
1.380	65	1	0.845	0.0412	0.764	0.926
2.610	60	1	0.831	0.0428	0.747	0.915
2.890	59	1	0.817	0.0444	0.730	0.904
3.420	57	1	0.802	0.0458	0.712	0.892
3.760	55	1	0.788	0.0473	0.695	0.880
4.170	52	1	0.773	0.0487	0.677	0.868
6.500	45	1	0.755	0.0506	0.656	0.855
7.190	42	1	0.737	0.0525	0.635	0.840
7.750	37	1	0.718	0.0547	0.610	0.825
7.900	36	1	0.698	0.0567	0.586	0.809
9.280	30	1	0.674	0.0594	0.558	0.791
9.440	29	1	0.651	0.0617	0.530	0.772
10.510	26	1	0.626	0.0642	0.500	0.752
10.530	25	1	0.601	0.0664	0.471	0.731
12.680	19	1	0.569	0.0700	0.432	0.707
13.040	17	1	0.536	0.0735	0.392	0.680
15.300	10	1	0.482	0.0834	0.319	0.646
16.380	8	1	0.422	0.0922	0.241	0.603
17.070	5	1	0.338	0.1056	0.131	0.545

18.950	1	1	0.000	NaN	NaN	NaN
--------	---	---	-------	-----	-----	-----

b)Intervalos de confiança para o estimador de Kaplan Meier com transformação Log-Log

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
0.030	78	1	0.987	0.0127	0.912	0.998
0.381	77	1	0.974	0.0179	0.901	0.994
0.410	76	1	0.962	0.0218	0.885	0.987
0.600	75	1	0.949	0.0250	0.869	0.980
0.770	74	3	0.910	0.0324	0.821	0.956
0.970	70	1	0.897	0.0344	0.805	0.947
0.980	69	1	0.884	0.0363	0.789	0.938
1.120	68	1	0.871	0.0380	0.774	0.929
1.320	66	1	0.858	0.0397	0.758	0.919
1.380	65	1	0.845	0.0412	0.743	0.909
2.610	60	1	0.831	0.0428	0.726	0.898
2.890	59	1	0.817	0.0444	0.710	0.887
3.420	57	1	0.802	0.0458	0.693	0.876
3.760	55	1	0.788	0.0473	0.677	0.864
4.170	52	1	0.773	0.0487	0.659	0.852
6.500	45	1	0.755	0.0506	0.639	0.839
7.190	42	1	0.737	0.0525	0.618	0.825
7.750	37	1	0.718	0.0547	0.594	0.809
7.900	36	1	0.698	0.0567	0.571	0.793
9.280	30	1	0.674	0.0594	0.543	0.776
9.440	29	1	0.651	0.0617	0.516	0.757
10.510	26	1	0.626	0.0642	0.487	0.737
10.530	25	1	0.601	0.0664	0.459	0.717
12.680	19	1	0.569	0.0700	0.421	0.693
13.040	17	1	0.536	0.0735	0.383	0.667
15.300	10	1	0.482	0.0834	0.313	0.632
16.380	8	1	0.422	0.0922	0.242	0.591
17.070	5	1	0.338	0.1056	0.148	0.539
18.950	1	1	0.000	NaN	NA	NA